



**SOSYAL AĞLARDA YAPAY ZEKÂ İLE VERİ ANALİZİ**

**Tohid JAVADI**

**YÜKSEK LİSANS TEZİ**

**BİLGİSAYAR MÜHENDİSLİĞİ ANA BİLİM DALI**

**GAZİ ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ**

**TEMMUZ 2019**

Tohid JAVADI tarafından hazırlanan “SOSYAL AĞLARDA YAPAY ZEKÂ İLE VERİ ANALİZİ” adlı tez çalışması aşağıdaki jüri tarafından OY BİRLİĞİ ile Gazi Üniversitesi Bilgisayar Mühendisliği Ana Bilim Dalında YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

**Danışman:** Doç. Dr. Necaattin BARIŞÇI

Bilgisayar Mühendisliği Ana Bilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum.

**Başkan:** Dr. Öğr. Üyesi H. Murat ÜNVER

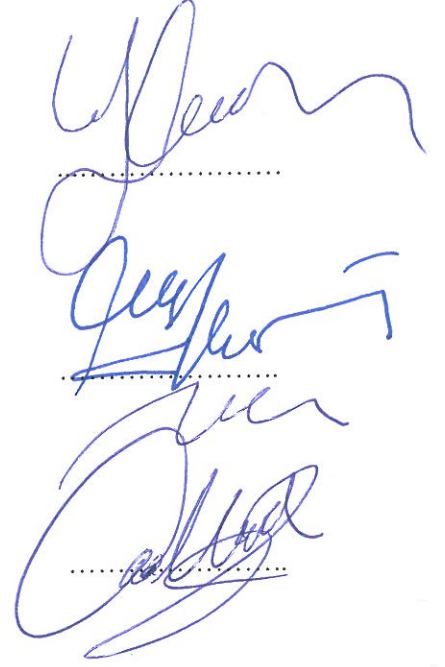
Bilgisayar Mühendisliği Ana Bilim Dalı, Kırıkkale Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum.

**Üye:** Dr. Öğr. Üyesi İsmail ATACAK

Bilgisayar Mühendisliği Ana Bilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum.



Tez Savunma Tarihi: 22/07/2019

Jüri tarafından kabul edilen bu tezin Yüksek Lisans Tezi olması için gerekli şartları yerine getirdiğini onaylıyorum.

.....  
Prof. Dr. Sena YAŞYERLİ

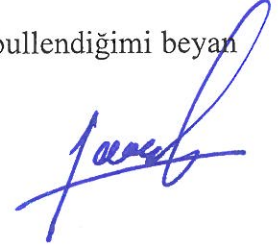
Fen Bilimleri Enstitüsü Müdürü

## ETİK BEYAN

Gazi Üniversitesi Fen Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmada;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmada yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu,

bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.



Tohid JAVADI

22/07/2019

SOSYAL AĞLARDA VE YAPAY ZEKÂ İLE VERİ ANALİZİ  
(Yüksek Lisans Tezi)

Tohid JAVADI

GAZİ ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ

Temmuz 2019

ÖZET

Bilim ve Teknoloji geçtiğimiz yüzyılın ikinci yarısından itibaren her geçen gün daha da artan bir ivme ile hızlanmaktadır. Bu gelişim, değişimi de beraberinde getirmiştir. Sosyal ağlarda büyük veri ve yapay zekâ teknikleri ile veri analizi isimli bu yüksek lisans tezinde “insanların olaylar karşısındaki duygu ve düşüncelerinin Twitter paylaşımlarından belirlenip belirlenemeyeceği üzerinedir. Makine öğrenimi, Yapay Zekâ ve önceden oluşturulmuş Corpus adı verilen büyük veriler (Big Data-BD) sayesinde %99.24 oranında doğrulukla sınıflandırması mümkün olmuştur. Bu sınıflandırma sayesinde hedef kitlenin ya da kişinin eğilimleri önceden tahmin edilebilmekte ve bu da politika yapıcıları ile tanıtım-pazarlama sektörü için yapılacak yatırımlar ve söylemler hakkında bir fikir verebilmeyi mümkün kılmaktadır.

Bilim Kodu : 92432  
Anahtar Kelimeler : Duygusal analiz, duygusal zekâ analizi, yapay zekâ, makine öğrenmesi, Twitter, Naive Bayes, N-Gram, karar ağacı, Python, Django  
Sayfa Adedi : 59  
Danışman : Doç. Dr. Necaattin BARIŞCI

ANALYSING DATA IN SOCIAL NETWORKS BY USING ARTIFICIAL  
INTELLIGENCE

(M. Sc. Thesis)

Tohid JAVADI

GAZİ UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

July 2019

ABSTRACT

Science and Technology has been becoming faster with an acceleration getting faster nowadays since the second half of the last century. This development resulted in change, too. At this work titled “Analyzing Data in Social Networks by Using Big Data and Artificial Intelligence” is on detecting people’s senses and thoughts by their shares in social media platform, Twitter. Thanks to Machine Learning (ML), Artificial Intelligence (AI) and Big data's naming PreSet Corpus, it could be possible to classify tweets with nearly 99.24% accuracy. With this classification, the trends of target one or group are predicted easily earlier and this makes possible taboo advice and strategy to make investment to politics-makers and marketing corporations.

Science Code : 92432

Key Words : Sentimental analysis, Twitter, Naive Bayes, machine learning, artificial intelligence, N-Gram, decision tree, Python, Django

Page Number : 59

Supervisor : Assoc. Prof. Dr. Necaattin BARIŞCI

## TEŐEKKÖR

Bu alıőmada bilgi ve birikimiyle bana yardımcı olan sayın hocam Do. Dr. Necaattin BARIŐCI ve yakın arkadaőım Tekin TOPUZ'a minnettarlıđımı sunarım ve Őukranlarımı bildiririm. Aynı zamanda bũyũk sabırla bana destek veren aileme ve eőim Nina SAVALAN'A teőekkũr ederim.



**İÇİNDEKİLER**

	<b>Sayfa</b>
ÖZET .....	iv
ABSTRACT.....	v
TEŞEKKÜR.....	vi
İÇİNDEKİLER .....	vii
ÇİZELGELERİN LİSTESİ.....	ix
ŞEKİLLERİN LİSTESİ.....	x
SİMGELER VE KISALTMALAR.....	xii
1. GİRİŞ .....	1
2. LİTERATÜR TARAMASI .....	5
3. MATERYAL VE YÖNTEMLER .....	9
3.1. Büyük Veriler.....	9
3.2. Veri Madenciliği .....	10
3.3. Makine Öğrenimi .....	11
3.4. Yapay Zekâ .....	15
3.5. Twitter .....	17
3.6. Python+Django.....	18
3.7. Tweepy .....	19
3.8. Postgre SQL .....	19
3.9. Google API.....	20
3.10. Javacript- JQuery.....	21
3.11. DataTable .....	21
3.12. N-Gram Analiz .....	22
3.13. Naive Bayes Algoritması.....	23
3.14. Karar Ağacı Algoritması .....	24



	<b>Sayfa</b>
3.15. Programın Çalışma Algoritması .....	25
3.16. Tweetlerin Temizlenmesi ve Corpusa Eklenmesi .....	26
3.17. Test ve Train Dataset Oluşturulması .....	27
3.18. Programın İş Akışı.....	28
<b>4. ARAŞTIRMA BULGULARI .....</b>	<b>31</b>
4.1. Naive Bayes Algoritması ve nltk.NaiveBayesClassifier .....	31
4.2. Karar Ağacı .....	33
4.3. Sklearn Classifier (BernoulliNB) .....	33
4.4. Uygulamanın Ekran Çıktıları .....	33
4.5. Canlı Tweet Analizi.....	43
4.6. Corpus Tweet Analizi.....	46
<b>5. DEĞERLENDİRME VE SONUÇ.....</b>	<b>51</b>
<b>KAYNAKLAR .....</b>	<b>55</b>
<b>ÖZGEÇMİŞ .....</b>	<b>59</b>

**ÇİZELGELERİN LİSTESİ**

<b>Çizelge</b>	<b>Sayfa</b>
Çizelge 4.1. Algoritmaların başarı oranı.....	49



## ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil 1.1. Global internet ve sosyal medya kullanım istatistikleri (2019).....	1
Şekil 3.1. İzlenmeyle öğrenme yapısı .....	13
Şekil 3.2. Denetimsiz öğrenmenin yapısı .....	14
Şekil 3.3. İki durumlu Markov Modeli .....	22
Şekil 3.4. Karar ağacı işlem mantığını gösteren örnek bir grafik .....	25
Şekil 3.5. Twitter API anahtarı .....	25
Şekil 3.6. Uygulama modelleri .....	26
Şekil 3.7. Tweetlerin temizlenmesi.....	26
Şekil 3.8. Yalın ve temiz tweet .....	27
Şekil 3.9. Programın iş akış şeması .....	28
Şekil 4.1. Corpusa ekle .....	32
Şekil 4.2. Giriş sayfası .....	33
Şekil 4.3. Ana yönetim paneli.....	34
Şekil 4.4. Naive bayes algoritması.....	35
Şekil 4.5. Karar ağacı algoritması.....	35
Şekil 4.6. N-gram .....	36
Şekil 4.7. Yalın tweet.....	36
Şekil 4.8. Yalın tweet.....	37
Şekil 4.9. Corpus-tweet.....	38
Şekil 4.10. Corpus-güncelle .....	38
Şekil 4.11. Corpus-analiz.....	39
Şekil 4.12. Corpus-analiz-tablo.....	39
Şekil 4.13. Canlı-analiz-tablo .....	40
Şekil 4.14. Canlı-analiz-tablo .....	41
Şekil 4.15. Canlı-tweet-akışı.....	42

<b>Şekil</b>	<b>Sayfa</b>
Şekil 4.16. Hashtag ekleme ekranı.....	42
Şekil 4.17. Canlı tweet analizi tablo-3 sınıflandırıcı birlikte .....	44
Şekil 4.18. Canlı tweet analizi tablo-naive bayes sınıflandırıcı.....	44
Şekil 4.19. Canlı tweet analizi tablo-karar ağacı sınıflandırıcı .....	45
Şekil 4.20. Canlı tweet analizi tablo-skylearn bernoulli (N-gram) classifier .....	45
Şekil 4.21. Canlı tweet analizi harita üzerinde gösterim .....	46
Şekil 4.22. Corpus tweet analizi- 3 sınıflandırıcı tek tablo.....	47
Şekil 4.23. Corpus tweet analizi-naive bayes sınıflandırıcı .....	47
Şekil 4.24. Corpus tweet analizi- karar ağacı sınıflandırıcı .....	48
Şekil 4.25. Corpus tweet analizi- skylearn bernoulli (n-gram) classifier .....	48
Şekil 4.26. Corpus tweet analizi-harita görüntüsü .....	49

## SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

### Kısaltmalar

### Açıklamalar

**AI**

Artificial Intelligence

**BD**

Big Data

**DA**

Duygusal Analiz

**ML**

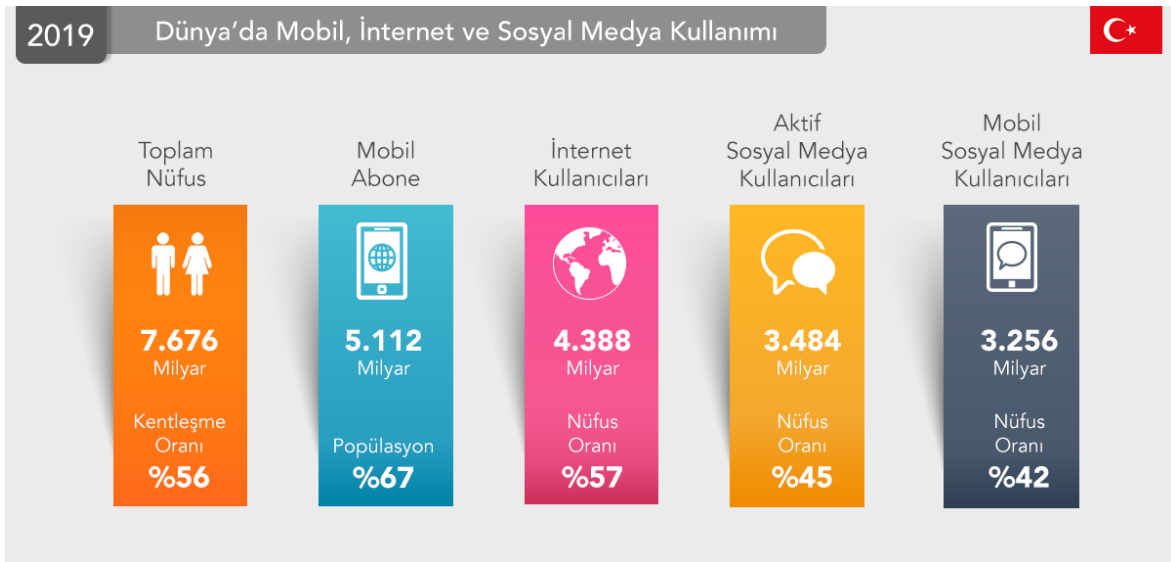
Machine Learning

**MÖ**

Makine Öğrenmesi

## 1. GİRİŞ

İnsanlar günümüzde eskiden hiç olmadığı kadar sosyallikten uzak dijital dünya içerisinde bu yoksunluklarını ya da ihtiyaçlarını giderme peşindedirler. Durum böyle olunca her gün sayıları ve içerikleri hızla artan sosyal medya mecraları ise daha da popüler hale gelmekte ve insanlar günlük yaşamdaki duygu, istek ve arzularını sosyal medyada siber dünyanın dili ile anlatmaya çalışmaktadırlar. Siber dünya her ne kadar sınırsız sayıda cihaz, altyapı ve kullanıcılardan oluşsa da genel anlamda insanlar çağlar boyunca edindiği bilgi, birikim ve kültürleri ile iletişim kurmaktadırlar. Bugün kullanılan yüzlerce sosyal medya araçlarında neredeyse kullanılan dil aynıdır. Bu dil kullanıcıların kendi ya da başka kullanıcıların paylaşmış oldukları içerikler hakkında yapmış olduğu yorumlar ya da beğeni ya da beğenmeme şeklinde karşımıza çıkmaktadır. Şekil 1.1’de görüldüğü gibi bu kullanım oranı her sene artmakta ve 2019 yılı itibariyle insanların %57’si internete bir şekilde erişim sağlamak ve %45’i de aktif bir şekilde sosyal medyayı kullanmaktadırlar. Dahası son dönemde meşhur olan duygu anlatımı için oluşturulmuş özel karakterler-emojiler kullanılmaktadır. Sosyal medya bugün itibariyle neredeyse dünya nüfusunun yarısına ulaşmış ve her yaşta her gruptan insanın eylemde bulunduğu, paylaşım yaptığı dijital bir platform durumundadır [1].



Şekil 1.1. Global internet ve sosyal medya kullanım istatistikleri (2019)

### Problem durumu / konunun tanımı

Çok sayıda kullanıcı web 2.0 etkisi ile yapılan paylaşımlara katkı sağlamakta ya da bizzat paylaşımı kendisi oluşturmaktadır. Bu kadar çok kişinin kullandığı ve bu denli yoğun içeriğin olduğu dijital platformdan anlamlı sonuçlar çıkarmak mümkün müdür? Bu paylaşımlardan elde edilebilecek anlamlı veriler ne işe yarar? Neredeyse rasgele web ortamına saçılan bu açık kaynak bilgiler kullanılarak bir çıkarımda bulunup, kişilerin gönüllü olarak kendilerinden iz bıraktıkları sanal ortamda onlar hakkında bilgi toplayıp işleme imkânına sahip olunabilir mi? Daha doğrusu kişilerin kendi istekleriyle paylaştıkları bilgiler kullanarak bir toplumun düzeni ve selameti için bu verileri akıllı sistemler vasıtasıyla kullanarak toplum daha yaşanabilir bir ortama dönüştürülebilir mi?

### Araştırmanın amacı

Kısaca Duygu (Sentiment) Analizi (DA denilen sosyal medya ortamlarındaki her türlü içeriğin makine öğrenimi (machine learning, MÖ) ve yapay zekâ algoritmaları (Artificial Intelligence - AI) ile işlenerek anlamlı sonuçlar elde etme sürecini ifade etmektedir. Günümüzün en çok tartışılan ve erişilen teknoloji çağının hızına ayak uydurmuş olan Twitter kitleleri en çok etkileyen bir alan haline gelmiştir. İşte “Duygusal (Sentiment) Twitter Analizi” isimli çalışma bu sosyal medya ortamı hedef alınarak yukarıda sorulan sorulara cevap ararken multidisipliner bir çalışma ile yapılan paylaşımlardan elde edilen anlamlı sonuçların nasıl ve ne şekilde kullanılacağı yönünde reklam ve pazarlamadan siyasete, spora ve toplumun hemen her alanı ile alakalı ilgili kişilere fikir verme ve genel eğilimleri (trendleri) gösterme amacı taşımaktadır. Bu çalışmada bilinen makine öğrenme ve yapay zekâ teknikleri kullanılarak olası “zararlı” kişiler ve kitleler hakkında tam isabetili bir tespitte bulunulmasa da, bir fikir edinilmesi için bir araç sağlamak amaçlanmaktadır.

### Araştırmanın önemi

Büyük verilerin bilgisayar sistemleri ile işlenerek anlamlı hale getirilmesi ilk başta incelenmesi oldukça zor olan bu verilerden anlamlı sonuçlar çıkarmayı mümkün kılacaktır. Dahası elde edilen bu sonuçlar kişilerin ve kurumların karar alma süreçlerine büyük katkı sağlayacaktır. Politika yapıcılar ve piyasa araştırmacıları elde edilen verilerle anlamlı ve doğru kararlar verebileceklerdir. Günümüzde devletler ve emniyet teşkilatları potansiyel

suçluları ve ileride soruna yol açabilecek kitleleri arařtırmak ve belirlemek için büyük miktarlarda kaynaklara ihtiya duyarlar. Örneđin bir toplumda řiddet yanlısı ve köktenci terör örgütlerine yakınlık ve sempati duyan bireyleri belirlemek ve onların olası eylemlerinin önüne geçmek için büyük miktarlarda para, zaman ve insan kaynađı ayırmak gerekmektedir. Hal bu ki bu kitlelerin çođu sosyal medyada ya kendi gerçek adlarıyla ya da sahte hesaplarla (ama gerçek konum ve teknik belirtilerle) faaliyet etmektedirler. Bu alıřma sayesinde otoriteler tarafından “zararlı” ve “istenmeyen” kitleleri sadece sosyal medyada yaptıkları faaliyetler sayesinde büyük dođruluk oranlarıyla tespit edebilir ve önleyici tedbirler alınabilir.

### Sınırlılıklar

Yapılan bu tezde kullanılan bilgisayar sistemlerinin verimsizliđi alıřmayı oldukça kısıtlamıřtır. Bunun yanı sıra Twitter řirketi kendi sistemi üzerinde geliřtiricilere atıđı API (Application Programming Interface) hizmetinde oldukça sınırlı sayıda tweete eriřim izni vermektedir. Söz konusu bu izin kiřinin attıđı son 200 tweeti kapsamaktadır. Sisteme eriřimde kullandıđımız API'nin belli zaman aralıklarında ve belli sayıda istekte bulunma hakkı vardır. Bu sebeplerden dolayı alıřma belli sayıda hedef kitleyi kapsayacak řekilde tasarlanmıřtır. Bu kısıtlama yalnızca corpus adı verilen süreç için söz konusudur. Kurulan sistem sayesinde anlık tweetlerin deđerlendirmesinde bir sakınca bulunmamaktadır. Ancak bir kullanıcının gemiř tweetleri bazında deđerlendirilmesi o kullanıcının son 200 tweeti temel alınarak yapılmıřtır. Diđer yandan proje geliřtirme esnasında birden fazla yazılım dili ve teknoloji denenmiřtir. Örnek olarak ilk bařta .NET teknolojisinin kullanımına karar verilmiř, ancak alıřmanın hazırlandıđı esnada Microsoft tarafından tam verimli yapay zeka kütüphanelerinin sunulmaması ve üçüncü parti kütüphanelerinde tam verimli alıřmamasından dolayı bu teknolojidenden vazgeilmiřtir. Bu sorunu ařmak adına çeřitli makine öđrenmesi ve yapay zekâ kütüphanesi olan Python dili kullanılmıřtır. Bu yazılım dili temel olarak seilse de web tabanlı uygulama olarak büyük sorunlar yařanmıřtır. Her ne kadar Django platformu bu dil için web tabanlı yazılım geliřtirmek için olanak sađlasa da pratikte web tabanlı yazılım süreci tahmin edilenden daha meřakkatli olmuřtur. Veri tabanı kısmındaysa aık kaynak ve ücretsiz (Postgre SQL) sürümlerle yol alınmıřtır. Bunun en büyük sebebiyse yüksek miktardaki lisans ücretleridir.



### Çalışma kapsamı

Bu çalışmanın ilk bölümünde kullanılan algoritmalar açıklanmıştır. İkinci bölümde bu algoritmalar hangi kütüphaneler kullanılarak ve web ortamında Python dili kullanılarak uygulandığına değinilmiştir. Üçüncü bölümde ise yapılan projenin işleyişiyle ilgili ekran çıktıları ve onlara ait açıklamalar yer almaktadır. Dördüncü ve son bölümde ise farklı 3 algoritmadan elde edilen verilerin kıyaslaması ve bulgular yer almıştır.



## 2. LİTERATÜR TARAMASI

Twitter'ın süper hızlı blog yapısından ve dünyanın her yerinden milyonlarca kullanıcısı sayesinde ürettiği anlık terabaytlık veri sayesinde büyük bir veri madenciliği alanı haline gelmiştir. Bu büyük yapıdan anlamlı sonuçlar çıkarmak isteyen birçok araştırmacı ve akademisyen bugüne kadar kayda değer on binlerce çalışma yapmışlardır. Bu çalışmada söz konusu literatürün birçoğundan faydalanılmıştır. Ancak yapılan araştırmaların sayısal olarak çokluğundan dolayı burada hepsine değinmek mümkün olmamıştır. Başlıca öne çıkan çalışmalara bakılacak olursa, "Google Akademik" arama motoru kullanılarak yapılan "Twitter ve Salgın Hastalıkları" ("Twitter epidemic research") başlığı altında 7000 akademik çalışma yapıldığı görülmektedir.

Bu alanda yapılmış kayda değer çalışmalardan biride Alexander Pak ve Patrick Paroubek tarafından Paris üniversitesinde 2010 senesinde yapılan gerçek zamanlı Twitter da "Duygusal Analiz" isimli çalışmasıdır [2]. Bu çalışmada insanların olumlu veya olumsuz duygularını tespit amaçlı bir çalışma yapılmıştır. Bu çalışmada ortalama her dakikada 108 olumlu, 75 olumsuz, 33 natürel, toplamda 216 adet tweet analiz edilebilmiştir.

Bu çalışmalardan en popülerleri Min Kang ve diğerleri tarafından New York'ta 2012-2013 (Influenza) grip salgını hakkında yapılan günlük Twitter geolocation analizleridir [3].

Kuşkusuz Twitter'ın tüm anti demokratik engelleri aşarak her türden kullanıcıya düşüncelerini açıklama ve söz sahibi olma hakkı vermektedir. Twitter sahip olduğu bu kapasite sayesinde insanların eğilimleri hakkında bilgi vermekte ve ortaya çıkan sonuçların bilimsel olarak analizi konusunda kapsamlı veriler ortaya koymaktadır. Google Scholar'da yapılan araştırmalarda dünyanın hemen hemen her yerinde yapılan seçimlerin sosyal medya üzerinde kazanılıp kazanılmadığı ortaya çıkmaktadır. 113 bin akademik çalışma arasından (Twitter election analysis GoogleScholar Search) Hao Wang ve diğerleri tarafından yapılan çalışmada 2012 yılında ABD'de yapılan Başkanlık seçimlerinin anlık veri akışı ve kullanıcı paylaşımları odaklı geolocation verileri de kullanılarak modellenmesi üzerine kurgulanmış ve politika yapıcılara ve politik aktörlere çok önemli bilgiler vermiştir [4].

Bu çalışmalarda İngilizce dili temel alınarak insanların attığı tweetler üzerinden dört siyasi rakibin oy potansiyeli ve muhtemel seçim sonuçları tahmin edilmeye ve yapılan seçimden

sonrada sonuçları analiz edilmeye çalışılmıştır. Bu çalışmada 4 başkanlık adayının taraftarlarının attığı tweetler toplanarak işlenmiş ve hangi aday hakkında ne kadar bahsedildiğine dair bir harita ortaya çıkmıştır. Ortaya çıkan haritayı ele alarak hangi adayın daha fazla kazanma şansının olduğu öngörülme çalışılmıştır.

İnsanların demokratik süreçlere katılımları ve insanlar arası ilişkileri kuvvetlendirmesi ile Twitter, son yıllarda toplumların iletişim kurma alışkanlıklarında büyük dönüşüm yaratmıştır. Etkisi hem toplumsal, hem ticari, hem de ekonomik yaşamda yoğun bir şekilde hissedilen Twitter, gün geçtikçe artan kullanıcısıyla da toplumun her kesiminde yaygınlaşmaya başlamıştır. Etkileşimlilik özelliği ve içeriğinin kullanıcılar tarafından belirlenmesi ile önemi giderek artan ağ, 2010'da yaşanan ve "Arap Baharı" olarak adlandırılan süreçte bilhassa toplumsal örgütlenme ve kitlesele eylemlerdeki etkisi ile de tüm dünyada önemli toplumsal olayların merkezinde yer almıştır. Böylece bir sosyal paylaşım ağı olmanın ilerisinde Twitter'a yüklenen anlam, toplumsal bir aktör niteliği kazanmıştır. Toplumsal olaylarda bir aktör olmasıyla birlikte Twitter, sosyal sorumluluk içeren etkinliklerde de önemli bir paya sahip olmuştur [5].

Etkileşim düzeyinin yüksek olduğu ve içeriklerinin kullanıcılarca belirlendiği Twitter, son zamanlarda demokratik katılım hususunda ve toplumsal olaylardaki etkisi ile önemli bir araca dönüşmüştür [6]. Twitter'ı sosyal ağlar içinde benzersiz ve önemli kılan ise, kaynak tarafından hedef gruplara en kolay ve en zahmetsiz şekilde mesajların iletilmesine imkân sağlamasıdır [7]. Twitter'ı önemli ve özel kılan bir başka özellikse, herhangi bir konuda kullanıcıların anahtar kelimeler ile aramalar yaparak, eş zamanlı olarak konuyla ilgili kullanıcılar tarafından gönderilen tweetlere erişerek bilgi sahibi olmalarına olanak sağlamasıdır [8]. Bu özellik sayesinde kullanıcılar dünyanın herhangi bir yerinde eş zamanlı olarak gerçekleşen olaylar hakkında bilgi sahibi olabilmektedirler. Bununla beraber Twitter'ın sayfasında bulunan (Trending Topic) bölümündeysel bölgesel olarak veya dünyada en çok nelerin konuşulduğuna ilişkin liste bulunmaktadır. Kullanıcılar bu listeyi inceleyerek güncel haberler hakkında bilgi sahibi olabilmektedirler. Trending Topic, kullanıcıların Hashtag (#) adı verdikleri ve kelimelerin başına geldiğinde linke dönüşen etiketler aracılığı ile bir konu etrafında tartışmasına dayanmaktadır. Konu belirli bir tartışma yüzdesine geldiğinde, Trendler listesine girmektedir.

Toplumların oldukça yoğun bir şekilde kullandığı bir mecranın etkili kullanılması, siyasal katılıma ve devlet-vatandaş ilişkisine önemli katkılar sağlamaktadır. Twitter siyasal örgütler bakımından örgütlenmenin yayılması ve seçmenlerin/vatandaşların siyasal iletişim sürecine, aktif olarak katılımlarının sağlanması anlamında uygun bir zemindir [9]. Son zamanlarda gerçekleştirilmiş olan çalışmaların ve araştırmaların birçoğunda, sanal ortamda toplumun siyasete ve karar alma süreçlerine katıldığı tespit edilmiştir [10].





### 3. MATERYAL VE YÖNTEMLER

Tezin ana çalışma prensibi olan Twitter paylaşımları odaklı kullanıcı eğilimlerini belirleme çalışması kapsamında Makine Öğrenimi (MÖ) ve Yapay Zekâ (YZ) kullanılarak oransal hesaplamalar ve algoritmalar sayesinde anlamlı sonuçlar çıkarılması amaçlanmıştır. Öncelikle bu kavramların ne olduğunu ve kullanılan yöntem ve kütüphaneler hakkında bilgi verilmiştir.

#### 3.1. Büyük Veriler

Şeker [11] veriyi, “Tanım itibariyle, herhangi bir işleme tabi tutulmadan, gözlem ve ölçüm yöntemleri ile ortamdan elde edilen her türlü değerdir” şeklinde tanımlamıştır. Diğer bir kaynağa göre veri, veri tabanında bulunan bilgi için kullanılan genel bir terimdir [12].

Yılmaz [13] ise veriyi, “tek başına anlam ifade etmeyen veya kullanılmayan, bununla birlikte enformasyona ve bilgiye temel oluşturan ilişkilendirilmeye, gruplandırılmaya, yorumlanmaya, anlamlandırılmaya ve analiz edilmeye gereksinim duyulan ham bilgi” olarak tanımlamıştır.

Veri, gözlemlerden, araştırmalardan, sosyal medyadan, internetten, algılayıcılardan vb. çok farklı ortamlardan elde edilen genel bir terimi ifade etmektedir. Veriler aşağıdaki gibi gruplara ayrılarak nitelendirilebilir [14]:

- Yapılandırılmış, Yapılandırılmamış, Yarı yapılandırılmış
- Güvenli / açık, özel / halka açık
- Statik, dinamik, akan
- Açık hükümet verisi
- Ücretli / ücretsiz
- Büyük veri
- Açık veri

Büyük Verinin ne olduğu konusunda birebir tanım olmasa da dünyanın önde gelen danışmanlık firmalarından Gartner tarafından yapılan bir değerlendirmede önemli sınırlar çizilmiştir [15, 16]. 3V adı verilen bu tanıma göre;

- 1- Dijital anlamda büyük hacim (Volume). En az 1 Petabyte ( $10^{15}$  bytes) boyutlarında,
- 2- Yüksek hızda veri akışı (Velocity).
- 3- Çeşitlilik (Variety)

büyük verilerin sınırlarını oluşturmaktadır.

Laney Gartner'ın bu tanımına benzer şekilde bu modelin başına Data kelimesi eklenerek 3D modeli olarak ileri sürmüştür [17]. Demchenko bir adım ileri giderek bu 3V-D3 tanımına Doğruluk (Veracity) ve Değer (Value) kavramlarını da ekleyerek 5V adını verdiği bir modelle Büyük Veriyi tanımlamıştır. Tanımlar bu şekildedir: Volume, Velocity, Variety, Veracity, and Value [18]. Değerlendirmeye göre bu tanımlar doğal süreç içerisinde gelişerek bugünkü kavrama hizmet edecek formata bürünmüştür. Yapısı ve hizmet ettiği ideal anlamda 5V tanımı oldukça yerinde bir kavramdır.

### 3.2. Veri Madenciliği

İnsanlığın geçmişten geleceğe tüm verileri günümüz teknoloji imkânları sayesinde dijital ortamda saklanmaktadır. Michael Lesk'e göre günlük 160 Terabaytlık veri üretilmektedir [19]. Önemli olan üretilen ve var olan bu büyük verinin anlamlı hale getirilmesidir. Aksi halde petabayt (1000 terabayt) boyutunda veriler sadece bir yığından ibaret olacaktır. Büyük veriler üzerinde yapılan bu anlamlandırma süreci beş (5) aşamadan oluşmaktadır. Bu aşamalar:

1. Aşama: Büyük verilerin toplanması ve veri ambarı adı verilen depolama birimlerine yüklenmesi.
2. Aşama: Veri ambarına iletilen verilerin ulaşılabilir ve üzerinde çalışabilir hale getirilmesi.
3. Aşama: Bu veriler üzerinde çalışacak profesyonellerin belirlenmesi ve verileri nasıl kullanacaklarının belirlenmesi.
4. Aşama: En uygun yazılımın seçilerek veriler üzerinde sıralama (sort) uygulanarak listelenen verinin elde edilmesi.
5. Aşama: Bu anlamlı verinin "an easy-to-share format" adı verilen kolay paylaşılabilir şekilde son kullanıcılara paylaşılmasıdır [20].

Veri madenciliği, bir veri analizi tekniği olup büyük miktardaki verinin analiz edilerek gizli kalmış bilgiye ulaşılmasını sağlamaktadır. Veri madenciliği, yeni bilgiler elde etme amacıyla çeşitli metotlardan oluşmakta olup çok farklı disiplinlerde kullanılmaktadır. Bilginin işlenerek verinin anlamlandırılması, yorumlanması ve ileriye dönük tahminler yapılması noktasında veri madenciliği yöntemleri fayda sağlamaktadır.

### 3.3. Makine Öğrenimi

İngilizce Machine Learning (ML) olarak adlandırılan bu kavram, insanın öğrenme ve öğrendikleri ile davranış göstermesi şeklindeki tutumlarının bilgisayar destekli bir cihaza yaptırılmasıdır. Burada öğrenen makinenin kendisidir. Bu öğrenme kesindir. Yani makine öğrenmeyi yaptıktan sonra beklenen kesin kez bu öğrenme yöntemi doğrultusunda faaliyet göstermesidir.

Makine Öğrenimi (MÖ) yaklaşımı, dilbilimsel ve/veya sözdizimsel özellikleri kullanan, normal bir metin sınıflandırma problemi olan DA'yı çözmek için kullanılan MÖ algoritmalarına dayanmaktadır. Bu yaklaşımda, popüler MÖ algoritmaları ve dilbilimsel özellikler uygulanmıştır. “Veri sözlüğü tabanlı yaklaşım daha önceden bilinen ve derlenmiş olan terimlerden oluşan bir duygu sözlüğüne dayanmaktadır” [21].

Samuel ve Webb “Encyclopedia of Machine Learning” isimli kitapta bu alanda yazılmış yüzlerce bilimsel araştırmayı bir araya getirmişlerdir. Peter Orbanz and Yee Whye Teh’e göre Makine öğrenimi eğitim veri kümesi (training dataset) adı verilen bir yapıdan uygun parametreler sayesinde öğrenme süreci olarak tanımlamaktadırlar [22].

İnsanlar karar verme süreçlerinde tecrübelerinden yararlanırlar. İlk defa karşılaşılan bir probleme nazaran daha önceden çözülmüş bir problemle karşılaşıldığında kolay bir karar verme süreci gerçekleşebilir. Bu örnekten yola çıkarak makine öğrenmesi algoritmaları insanların tecrübeleriyle açıklanabilir. İnsanların daha önceden edindiği tecrübe öğrenme veri kümesi olarak tanımlanabilir. Makine öğrenmesi algoritmaları çözülmesi istenilen sorunla ilgili bu veri kümesinden tecrübe kazanır ve daha sonra bu tecrübelerini başka problemleri çözmek için kullanırlar. Makinelerin karar verme sürecinde daha iyi bir konuma gelebilmesi için daha çok tecrübeye yani daha büyük hacimli bir veri kümesine ihtiyaç duyulmaktadır.



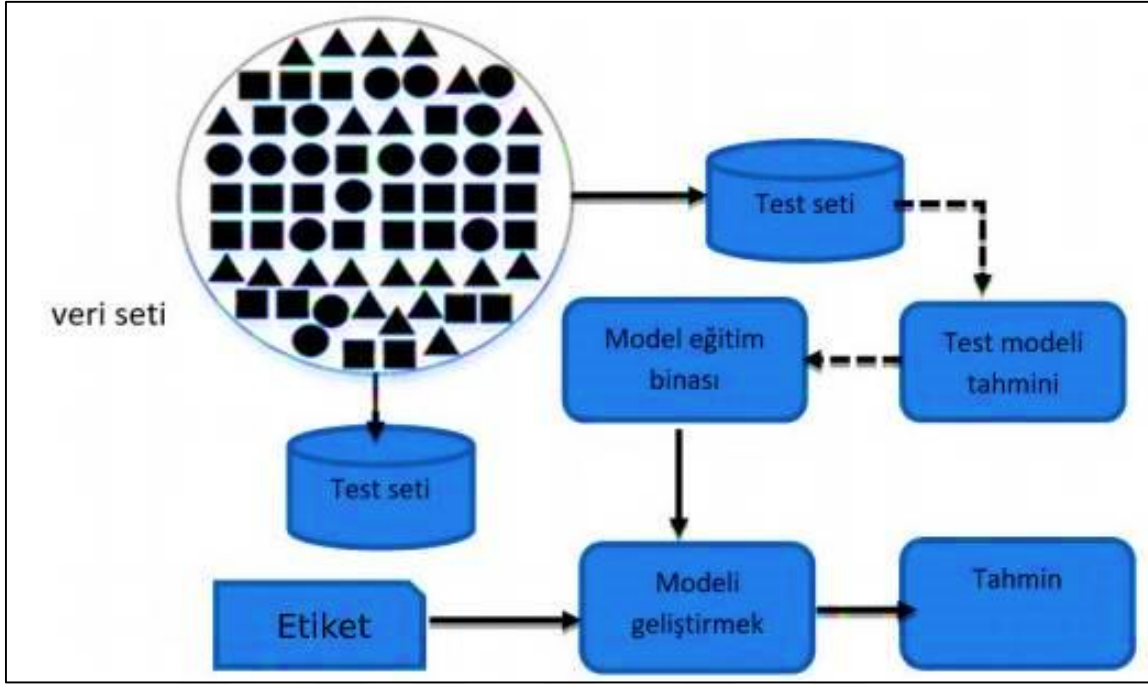
Makine öğrenmesi algoritmalarının bir kısmı öğrenme süreçlerinde veri kümesi içerisinde veriye özgü özelliklerin çıkarımını kendisi gerçekleştirirken bir kısmı ise bu özellikleri algoritmayı kullanan kullanıcıdan istemektedir. Veriye özgü özellikler çıkarılarak bir vektöre dönüştürülür. Bu vektör “özellik vektörü” olarak adlandırılır. Özellik vektörü oluşturmak için; bilgi kazanımı, olasılık oranı, log olasılık oranı ki kare, terim frekansı ve ters doküman frekansı gibi birçok farklı yöntem mevcuttur [23].

Bilinen birçok farklı makine öğrenme algoritması mevcuttur. Bu farklılık algoritmaların problemi değişik bakış açıları ile ele alınmasından kaynaklanmaktadır. Makine öğrenmesi algoritmaları özellik vektöründen edinmiş olduğu bilgileri kullanarak modeller üretmektedir. Üretilen bu modeller daha sonra sınamaya verisi ile sınanmaktadır. Elde edilen sonuçlardan modellerin performansı değerlendirilmekte ve problemlerin çözümünde en iyi olan model belirlenmektedir. Belirlenen bu model ilgili problemin çözümünde kullanılmaktadır.

Makine öğrenimi iyi hazırlanmış büyük bir eğitim kümesi gerektirmektedir. Böylece sistem neyin ne olduğunu anlayabilecek yeterliliğe gelmektedir.

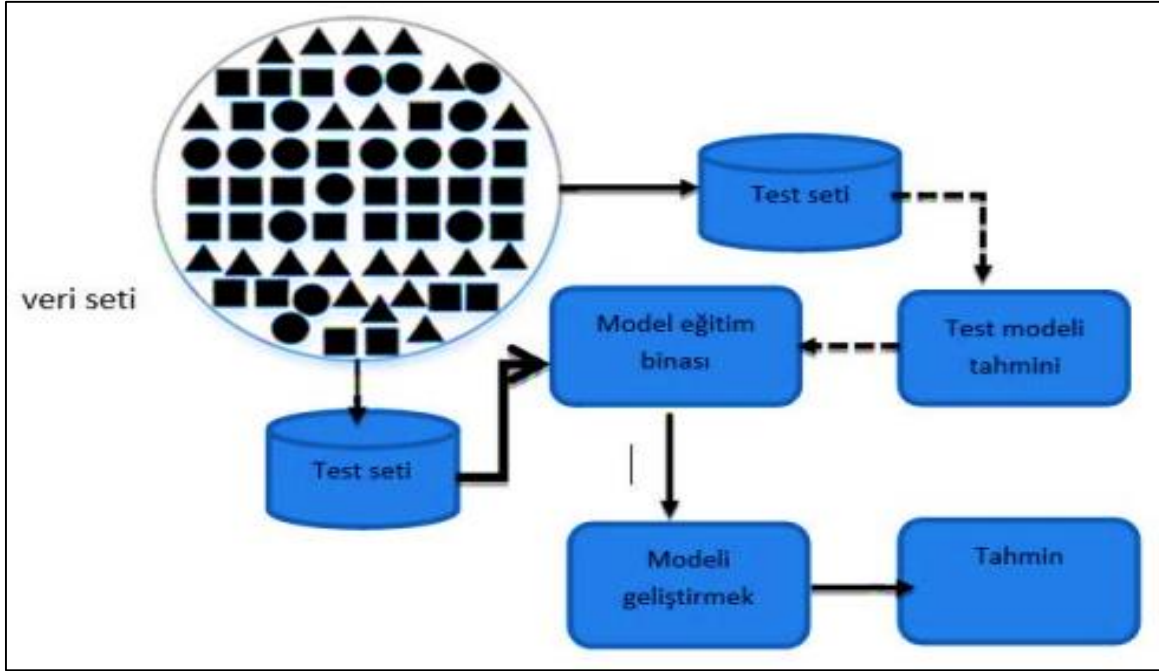
Makine öğrenme algoritmalarının iki aşaması söz konusudur [24]. İlk adım eğitimidir ve bu adımda öğrenim gerçekleşmekte, eğitim örnekleri incelenmekte ve buna dayalı bir sınıflandırma modeli yapılmaktadır. İkinci aşama ise bir testtir ve eğitim aşamasında yapılan bir model yeni gözlemlenen örnekleri sınıflandırmak için kullanılmaktadır. Makine öğrenme algoritmaları dört ana türe ayrılmıştır.

Denetimli öğrenme yönteminde, algoritmayı öğretmek için örnekler etiketlenerek kullanılmaktadır [25]. Etiketli örnekler, verilerin sonuçları ve yanıtları ile birlikte mevcut olduğu anlamına gelmektedir. Örneğin, öğrenme algoritması öğrenme aşamasında hastanın vakasını sağlıklılarından nasıl ayırt edilebilir diye öğretir, örnekler eğitim için etiketlenirler. Bir öğrenme algoritması sağlıklı veya bir hastalığı olanları nasıl sınıflandırması gerektiğini öğretmektedir. Şekil 3.1 de görüldüğü gibi eğitimden sonra algoritma, yeni bir numunenin sağlıklı olup olmadığını belirlemek için etiketlenmemiş yeni verileri sınıflandırabilir. Karmaşık konuları denetimli öğrenmek performans açısından daha iyi olabilmektedir.



Şekil 3.1. İzlenmeyle öğrenme yapısı

Denetimsiz öğrenmede, hedef değişken olarak bilinmemektedir ve algoritmanın çıktısı belirsizdir [26]. Örneğin, müşterileri kümelendirmek için bir mağaza, kişisel bilgiler ve müşteri satın alımları temelinde otomatik olarak eşit gruplara bölünebilir. Denetimsiz öğrenmede, istatistiklere dayanan tahminler, etiketlenmemiş örnekler arasındaki kalıplar, ilgileşimleri ve ilişkileri tanımlamak için kullanılmaktadır. Kalıplar tanımlandığında, Şekil 3.2'de görüldüğü üzere algoritma veri kümesindeki sınırları tanımlamak için istatistikler kullanır. Aynı modeldeki örnekler bir grupta kümeleşirler. Veri kümeleme işlemine devam ederek, algoritma veri kümesi modelini tanır ve yeni veriler için tahmin yapar. Makine istatistik ve kümelenmesinde kümelenme analizi, denetimsiz öğrenme dallarından biridir. Birbirine benzer örnekler gruplara ayrılırlar, bu gruplar küme olarak adlandırılır. Bu nedenle, kümeleme, örneklerin benzer olduğu bir koleksiyonudur ve diğer kümelerdeki örneklerle uyumlu değildir. Aslında kümelenme, önceden tanımlanmış sınıfların bulunmadığı denetimsiz bir sınıftır. Başka bir deyişle, kümelenme bir tür denetimsiz öğrenme türüdür.



Şekil 3.2. Denetimsiz öğrenmenin yapısı

Güçlendirilmiş öğrenmede, belirli bir kararı almak için bir makineye kontrolör programı öğretilmektedir [27]. Makine mevcut pozisyona (mevcut değişkenler kümesi) ve yetkili işlemlere (örneğin ileriye veya geriye doğru hareket ederek) dayalı bir karar alır. İlk seferde bu karar tamamen rastgele olabilir ve ortaya çıkan her eylem veya davranış için, sistem ona geri bildirim verir, bu geri bildirimlerde, makine doğru kararı vermiş olup olmadığını öğrenir. Güçlendirilmiş (takviye) öğrenimi interaktif bir yöntemdir ve denetimli öğrenme yönteminden farklıdır. Denetimli öğrenmede yedek vektör makinesi veya karar ağacı gibi, sınıf etiketleri eğitim aşamasında belirlenir. Fakat Güçlendirilmiş (takviye) öğrenmede, etken kendi öğrenmesidir. Öğrenme ne kümelenme ne de sınıflandırmadır, belki bir öğrenme metodudur [28]. Tüm denetimli öğrenme algoritmalarında, bir eğitim aşamasına sahipler ve modeli, eğitim ilk örnek (eğitici şablonlarla) yardımıyla oluşturulur ve sonra öğrenilene göre bir test aşaması yapılır ve test verileri kategorize edilmektedir. Güçlendirilmiş öğrenim, karar vermek için önemli olan tüm konular ve oyunlar için (satranç ve yönlendirme gibi) kullanılmaktadır. Böylelikle, etkenler başladıkları noktadan nasıl hareket edebileceklerini, hedeflerine nasıl ulaşacaklarını ve her durumda en iyi uygulamanın hangisi olduğunu öğrenmektedirler. Bu yöntem arayışı interaktif bir yöntemdir ve öğrenim deneme ve yanılma ile eklenir, yani öğrenen kişi, çevre ile etkileşim konusundaki bilgisini bir deneme ve hata olarak öğrenmektedir. Güçlendirilmiş öğrenme olasılıksaldır. Etken, her eylem için bir ödül

alır. Ödül ne kadar iyi olursa, iş o kadar iyi olur. En büyük hedef en fazla ödül elde etmektir, yani hedef en yüksek ödül olmalıdır.

Yarı-denetleme öğretim aynı zamanda bir öğrenme türüdür ve öğrenmenin doğruluğunu artırmak için, hem sınıflandırılmış (etiketli) hem de sınıflandırılmamış (etiketlenmemiş) verileri eşzamanlı olarak kullanmaktadır [29]. Yarı-denetleme, öğrenme-öğretici olmak üzere iki aşamalı sınıflandırma ve kümeleme içermektedir. Kümeleme algoritması, etiketli numunelerden ve etiketlenmemiş örneklerden oluşan eğitim örneklerini benzer kümelere ayırmaktadır. Eğitim örneklerinin kümeleneşinden sonra, küme örnekleri, numuneleri eşlemek için (numuneleri izole etmek) etiketli örneklerden küme örneklerini tanımlamak için kullanılmaktadır. Test aşamasında, eğitim dışı ve etiketlenmemiş örnekler kullanılmaktadır.

### 3.4. Yapay Zekâ

Yapay Zekâ genel anlamda insanlar tarafından gerçekleştirilen mantıksal süreçleri de içine alan düşünsel boyutta ikilemlerin bilgisayarlarca yaptırılması yerine getirilmesini disiplin edilmiş bir bilim olarak tanımlanabilir. Alan Turing'in çalışmaları üzerine çokça akademik yayın ortaya koyan Copeland de benzer bir şekilde tanım yapmaktadır [30].

“Yapay zekâ zeki makineler, özellikle zeki bilgisayar programları yapma mühendisliği ya da bilimidir.” “Düşünme, anlama, faaliyete geçirmeyi sağlayacak bilgi işleme çalışmasıdır.” “Yapay zekâ bağımsız makineler -bu makineler insan olmaksızın karmaşık işler yapabilir- inşa etmek için araştırma yapan bilişsel bilim dalıdır. Bu hedef makinelerin düşünmesini ve anlamasını gerektirir. Bu konuda akıl almaz ilerlemeler sağlanmışsa da hedefe bakıldığında hayal gibi gözükmektedir.” [31].

Yapay Zekânın gelişim aşaması incelendiğinde beş önemli evre söz konusudur:

1. Çağımızdaki bilgisayarı ilk olarak tanımayan Alan Turing, 1950'de makine ve insan etkileşimini hedef alan ve makinenin inandırıcılık seviyesini ölçmek için bir sistem geliştirmiştir. Adı Turing Testi olarak bilinen bu çalışma temel olarak bir makinenin insan gibi düşünüp düşünemeyeceği ve etkileşim içinde olduğu kişiyi inandırıp inandıramayacağı sorunsalı üzerine bina edilmiştir.

2. 1956'da "Yapay Zekâ" kavramı ilk defa Bilgisayar Bilimcisi John McCarthy tarafından Dartmouth Üniversitesi'nde bir konferansta kullanılmıştır.
3. 1973 yılında Prof. Sir James Lighthill, makinelerin satrançta "deneyimli bir amatör" seviyesinden asla daha fazla bir şey elde edemeyeceğini savunmuştur.
4. 1997'de IBM'in süper bilgisayarlarından Deep Blue, dünya satranç şampiyonu olan Garry Kasparov'u yenmiştir. Deep Blue, kapasite olarak saniyede 200 milyon olasılığı analiz edebilmiştir.
6. 2016 yılında Google ekibi, bilgisayarlarını birbirinden şifreleyebilecekleri ve gizleyebilecekleri tarafsız ağlar oluşturarak bilgisayarlarını gizli tutmayı öğretmişlerdir.

Yapay Zekânın Üstün Yönleri aşağıdaki gibi açıklayabiliriz;

- Kalıcıdır: İnsanlar öğrendikleri ve deneyimledikleri olayları ve konuları, benzerlerini yaşamadıkları sürece zamanla unutabilirler. Fakat tekrarlanan ve insanlar için önemli tecrübelerle dair bilgiler bellekte sürekli olarak tutulmaktadır. İnsanların unutma hızı, konuya verilen önem derecesi ile ters orantılıdır. Bir diğer ifadeyle konuya verilen önem azaldıkça, unutma hızı artar [32]. Unutma sırasında, önce detaylar, ardından ana bilgiler kaybolmaktadır. Unutulan bilgiler, belirli hatırlatmalarla yeniden kazanılabilmektedir. Unutma özelliği, insanın belleğinin lüzumsuz verilerle işgal edilmesini önlemektedir. Bilgisayarların belleklerine yazılmış olan veriler bilinçli olarak silinmediği müddetçe, bellekte kalmaktadır. Başka bir deyişle bilgisayarların unutma yeteneği yoktur.
- Paylaşılabilir: İnsan zekâsı, eğitim, öğrenme ve çevre etkenleri ile gelişmektedir. Bu birikimlerin diğer bir insana transferi, usta çırak ilişkisiyle mümkün olabilmektedir ve transfer işi uzun zaman almaktadır. Bir insanda gelişen zekâ tamamen bir başka insana aktarılamaz. Fakat yapay zekâ bir bilgisayarda oluşturulduğu için bunun bir başka bilgisayara transferi, kopyalama yöntemi ile kolaylıkla yapılabilir.
- Kolay elde edilebilir: Bir bilgisayarın zekâ seviyesinin yükseltilmesi, bir insanın zekâ seviyesinin yükseltilmesine oranla daha kolaydır; kolay olduğundan maliyeti de daha azdır.
- Tutarlıdır: İnsandan insana aynı olay karşısında verilecek tepkiler değişiklik gösterir. Hatta benzer olan iki olayda bile, aynı insanın tepkileri farklı olabilmektedir. Bunun sebebi insanların zekâ düzeylerinin farklı olması ve insan olmalarıdır. Yapay zekâda aynı veya benzer olaylarda verilen tepki her defasında aynıdır. Farklı zekâ düzeyindeki

yapay zekâ sistemlerinden aynı olay karşısında farklı tepkilerin görülmesi, zekâ düzeyiyle ilgilidir.

- Kayıtlıdır: Bir olay veya durum karşısında yapay zekânın ürettiği tepkiler kaydedilmektedir. Dolayısıyla, benzer bir durumla karşılaşıldığı zaman eski tepkiler gözden geçirilmekte ve en yakın olan tepki gösterilmektedir. Ancak insanlar, karşılaştıkları her olayda, geçmişte yaşadıkları benzer olayı akıllarına getirmekte veya onu hiç düşünmeden yeni bir tepki gösterebilmektedirler.

### 3.5. Twitter

Twitter Jack Dorsey tarafından 2006 yılında geliştirilmesinden bu yana popülaritesini her geçen gün artırmış ve içerdiği uygulamalar, programlama ara yüzü, kısa mesaj gönderimi ve alımı konusundaki olanakları ile internet dünyasının SMS'i olarak anılmaya başlamıştır. Twitter, bir sosyal ağ ve mikro blog sitesidir. Twitter, kullanıcılarına Türkçede “cıvılda” anlamında gelen tweet adı verilen en fazla 140 karakterlik metin yazmaya imkân veren yeni nesil iletişim aracı olmuştur. Kullanıcılarca gönderilen tweetler herkes tarafından görülebilir. Fakat kullanıcılar gönderdikleri tweetleri yalnızca kendi takipçilerinin görebileceği şekilde sınırlandırabilmektedirler. Kullanıcılar üye oldukları kişilerin veya kurumların takipçi sayısını ve bunların kimleri takip ettiklerini görebilmektedirler. Bütün kullanıcılar Twitter'ın web sitesi aracılığı ile tweetlerini gönderebilirler ve alabilirler. Kullanıcılar Twitter'ın web sitesine uyumlu harici uygulamalar (tablet, akıllı telefon vb.) veya belli ülkelerde erişilebilen kısa mesaj servisleri tarafından giriş gerçekleştirebilirler [33].

Twitter Amerika merkezli yeni nesil bir haberleşme ve sosyal medya hizmeti veren bir şirkettir. Kullanıcılar mesajlarını 140 karakter sınırlandırması ile paylaşabilmektedir. Bu mesajlara “tweet” denilmektedir. 7 Kasım 2017’de aldığı kararla bu mesaj karakter sınırını 2 katına (280 karakter) çıkarılmıştır. Bu güncellemede Çin, Japonya ve Kore kapsam dışında kalmıştır. Kayıtlı kullanıcılar mesaj atabilmekte, zaman çizelgesi adı verilen akışta veya herhangi bir başkasının tweetini retweet ve beğeni yapabilmektedir. Twitter’a kullanıcılar “Twitter.com” internet sitesinden veya IOS, Android ya da başka platformlar (Samsung TV vb..) için geliştirilmiş uygulamalar aracılığı ile sisteme girebilirler. Twitter geliştiriciler ve özel amaçlı veri ihtiyacı olan kişi ya da kurumlar için “Twitter api” adı verilen bir sistem aracılığı ile veri hizmeti sağlamaktadırlar [34].

### 3.6. Python+Django

Python (<https://www.python.org/>), yüksek seviyeli (high-level), genel amaçlı (general-purpose) ve yordamlanabilir (interpreted) bir programlama dilidir. Bir derleyiciye ihtiyaç duymayıp, kendi yorumlayıcısı ile çalışmaktadır. Guido van Rossum tarafından 1991'da yayınlamıştır. Guido'nun felsefesi kodun okunabilirliği ve indentation denilen 4 karakterlik girinti idi. Günümüzde Python hemen hemen tüm alanlarda (masaüstü, web, grafik, oyun vb.) kullanılmaktadır.

Bu çalışmada ilk başta .NET teknolojisi kullanılmak istenmiştir yalnız bu teknolojisinin kütüphane yetersizliğinden dolayı Python dili tercih edilmiştir. Python dili geniş geliştirici kitlesine sahip olmasından dolayı yapay zekâ ve makine öğrenmesi konusunda büyük olanaklar sağlamaktadır. Sadece makine öğrenme alanında onlarca açık kaynak kütüphane mevcuttur. Bu açık kaynakları kullanarak çok büyük oranda istenilen çalışma yapılabilmektedir. Tabii ki Python dili de kendine özgü zorlukları vardır. Bu çalışmada ortaya çıkan en büyük zorluklardan biri Python'a özel ve ücretsiz olan bir geliştirme aracı (IDE) olmuştur. Bu çalışmada çeşitli geliştirme ortamları denenmiş ve en son Pycharm geliştirme ortamı seçilmiştir.

Django; Python kodları ile hazırlanmış bir web ara yüzüdür. Yani Php ile hazırlanan OpenCart, Wordpress, NukePHP gibi Django'da Python kodları ile geliştirilmiş bir araçtır [35].

Django (<https://www.djangoproject.com/>), Python temelli ücretsiz ve açık kaynak kodlu bir web-framework (çerçeve)'dir. Güncel yazılım pattern (örüntülerinden) Model-View-Template (MVT-ModelGörüntüKalıp) sistemi üzerine kurgulanmıştır.

Bu çalışmada Django geliştirme aracının 2.1 sürümü kullanılmıştır. Django genel olarak Python web uygulamaları için geliştirme sürecinde rahatlık sağlasa da Framework bağımlı olmasından dolayı yeni güncelleme geldiğinde büyük zorluklar çıkarmaktadır. Özellikle hata giderme, hata bulma ve akıllı öneriler araçları eksikliğinden dolayı geliştirme süreci normalden fazla sürmüştür.

### 3.7. Tweepy

Tweepy, Python programcılar için geliştirilmiş, kolay kullanımı ile ön plana çıkan TwitterApi'ye erişimi sağlayan bir Python paketi-modülüdür [36, 37]. Yazılımla Twitter'a bağlantı sağlamak ve Twitter'dan istekte bulunarak veri çekmek amacı ile Python programlama diliyle uyumlu olan ve Twitter API bağlantısını sağlamak için Tweepy kullanılmıştır [38]. Bu Modül, Python'da Twitter için geliştirmeler yapan bir kütüphanedir. Twitter API'ı kullanılarak yazılmıştır. Tweepy modülü, Python yazılım diliyle beraber kullanılarak Twitter'ın geliştiriciler için sunmuş olduğu API hizmetine erişim sağlamaktadır [39]. Bu çalışmada, Twitter üzerinden veri çekme işlemleri yapılırken kullanılmıştır.

### 3.8. Postgre SQL

“Dünyanın en ileri seviyeli ilişkisel veri tabanı” sloganı ile reklamını yapan, birçok geliştirici ve profesyonel iş dünyasının tercihi olan başarılı bir ilişkisel veri tabanı sunucusudur [40].

PostgreSQL, açık kaynak kodlu, platform bağımsız gelişmiş bir nesne ilişkisel (ORDBMS) veri tabanı yönetim sistemidir. Modern kurumsal veri tabanı özellik ve kabiliyetine sahiptir. PostgreSQL'in, 1977'de başlamış olan 20 yılı akademik olmak üzere, son 20 yılı endüstride geçen 40 yıllık bir geçmişe sahip olan en eski açık kaynak kodlu yazılımlardan birisidir. PostgreSQL, tüm dünyada kamuda CERN, NASA, Fransa, İngiltere, G. Kore, vb. önemli devlet hizmetleri sunan uygulama sistemlerinde, finans ve Telekom sektörlerinde iş kritik uygulamalarda, Apple, Microsoft, IBM, Amazon, vb. dünyada önde gelen üreticilerin ürünlerinde, araştırma merkezlerinde ve üniversitelerde, küçük ölçekli projelerden çok büyük ölçekli kurumsal altyapılarda güvenilerek kullanılmaktadır. PostgreSQL'in, önde gelen ticari veri tabanı ürünleriyle rekabet edecek kurumsal veri tabanı özelliklerinin yanında günümüzdeki dijital dönüşüm projeleri ve teknolojileriyle uyumlu pek çok yeni ve yenilikçi özellikleri vardır (Örneğin; paralel sorgular, dizi şeklindeki veri tipleri, JSON veri tipini desteklemesi ve üzerinde sorgu çalıştırabilmesi) [41].

PostgreSQL, diğer birçok gelişmiş veri tabanında olduğu gibi dışarıdan kendine bağlanan istemcilerle belli bir dilde konuşup anlaşabilmek için bir İstemci/Sunucu Protokolüne sahiptir. “Herhangi bir istemci bu mimarinin getirdiği standartlar doğrultusunda veri tabanı üzerinde protokolce desteklenen tüm sorgulama işlemlerini gerçekleştirebilir. Bunun için yapılması



gereken tek şey veri tabanı sunucusu ile kurulacak bir soket bağlantısı üzerinden basit read/write sistem çağrılarında bulunmaktan ibarettir. Fakat hiçbir kullanıcı - özel amaçlar doğrultusunda sıfırdan bir kütüphane geliştirmedeği sürece - PostgreSQL'in bu iç işleyişi ile uğraşmak istemez. İşte bu noktada devreye API (Application Programming Interface/Uygulama Programlama Ara yüzü) kavramı girmektedir. PostgreSQL, dışarıdan kendisine bağlanacak programlama dilleri için bir ara katman oluşturan uygulama ara yüzlerine sahiptir. Bu sayede, veri tabanına ulaşmak isteyen programcılar, PostgreSQL'in kendi iç işleyişi ile ilgili hiçbir çalışma mekanizmasından haberdar olmak zorunda kalmadan da, kullandıkları dilin kendilerine sundukları API kütüphanesinden faydalanarak kolaylıkla veri tabanı ile etkileşime geçip, ilgili sorgulamalarını gerçekleştirebilirler. Yazılım piyasasındaki neredeyse tüm popüler programlama dilleri PostgreSQL veri tabanı için bir programlama ara yüzünü beraberinde sunmaktadır” [42].

### **3.9. Google API**

Google sahip olduğu teknolojik kapasiteyi geliştiriciler ve iş dünyası için kullanılabilir hale getirmiştir. Google hesabından oluşturulan API anahtarları sayesinde Google haritalar eklentisi özel amaçlı kullanılabilir [43].

Google ekiplerinin geliştirdiği Web Haritalama Servisi'ne (<http://maps.google.com/>) tek düzen kaynak konumundan ücretsiz olarak erişilmektedir. Pek çok haritalama servisini de destekleyen Google Maps, ticari kullanıma açık değildir. Google Ride Finder ve Google Transit'i bünyesinde bulunduran bir web sitesi ile hizmete sunulan Google Maps web haritalama servisinin; yol ve yer bulma, konum belirleme, yerleşik web kameraları görüntülerini iletme gibi birçok işlevi vardır. Wikipedia ile bilgiye erişimi kuvvetlendiren Google Maps, fotoğraf servisleri ile de geliştirilmeye açık bir bilgi ortam sunmaktadır. Bunun dışında Application Programming Interface (API) desteğiyle 3. parti web siteleri de desteklenmektedir. Google Maps'te 3D özelliği de bulunmakta, bu hizmetten yararlanabilmek için tarayıcıya eklenti kurulması gerekmektedir [44].

Google Maps, var olan imkânlarını erişimcileri ile paylaşmakta, onlarla işbirliği içerisinde bulunarak veri tabanını her geçen gün genişletmekte, kullanıcılarına kendi uygulamalarını geliştirme imkânı tanımaktadır. Bunun dışında sektörde faaliyette bulunan çeşitli şirketler ile yaptığı anlaşmalarla kullanıcılarına o şirketlerin imkânlarından yararlanma olanağı da

sağlamaktadır. Google Maps; 3D bina çizimi, video ve fotoğraf yükleme, kullanıcılar tarafından çizilen projelerin özel olarak kaydedilmesi ve paylaşılabilmesi, üniversitelerin ve eğitimcilerin akademik ve eğitim çalışmalarına yönelik eklentileri, gönüllü kullanıcılarının çalışmaları gibi özellikleri ile geliştirilebilirlik kıstası açısından son derece başarılıdır.

### 3.10. Javacript- JQuery

Javascript(JS) yüksek seviyeli yorumlanabilen bir programlama dilidir. JQuery ise JS temelli bir kütüphanedir.

JQuery etkileyici görsel bileşenler ve etkiler yapmak amacı ile oluşturulmuştur. Bunların çoğu ana JQuery kütüphanesinde mevcuttur ancak JQuery grubu daha yüksek seviyeli yapıları ayırıp ayrı bir pakette toplayarak bunu düzenli bir kütüphane haline getirip JQuery yapısının en üst katmanına yerleştirmişlerdir. Bu katmana (kütüphaneye) JQuery UI (Jquery Kullanıcı Ara yüzü) adı verilmektedir [45].

JQuery'de istenilen eklentiler oluşturulup bunlar bütün JQuery projelerinde yeniden kullanılabilir ve diğer program geliştiren kişiler ile rahatlıkla paylaşılabilir. Bununla birlikte JQuery, var olan HTML kodu karmaşası ile sınırlanamaz. Diğer bir ifade ile kullanışlı fonksiyonlar aracılığıyla yeni sayfa elemanları veya doküman parçaları eklenebilir. Sayfanın herhangi bir yerine yeni HTML parçası eklemeleri yapılabilir. Bunun dışında herhangi bir şeyi silme, çıkarma işlemleri de rahatlıkla yapılabilir.

### 3.11. DataTable

DataTable, HTML Tabloların interaktif şekilde kolay ve hızlı bir şekilde işlenmesini sağlayan bir JQuery kütüphanesidir.

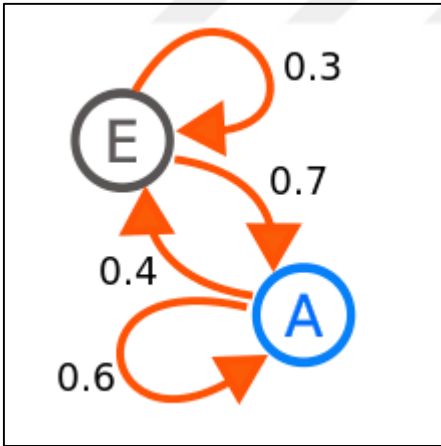
DataTables ile tablolarda aşağıdakiler yapılabilir [46];

- Sayfalama, eşzamanlı arama ve çok sütunlu sıralama
- Hemen hemen her veri kaynağını destekleme
- DataTables, JQuery UI, Bootstrap, Foundation tablo temalarını kullanma imkânı
- Tamamen çeviri yapılabilir,
- Ücretsiz açık kaynak (MIT lisansı) olması

- Sütun genişliklerini verimli kullanılması
- Gizli sütunlar
- Tabloların dinamik oluşturulması
- Ajax ile otomatik veri yüklenmesi
- Tek veya çoklu sütunda arama desteği olması
- Çok geniş eklenti desteği olması
- Kapsamlı dokümantasyonu olması ve devamlı geliştiriliyor olması.

### 3.12. N-Gram Analiz

N-gram analizi istatistiksel dil modelleme amacıyla kullanılan yöntemlerden birisidir. Gram adı verilen harf ya da kelimenin bir sonraki elemanla ilişkisine dallandırılarak bir sonraki elamanın ne olduğu konusunda bilgi vermeyi amaçlayan bir yöntemdir [47]. Şekil 3.3’de görülüşü gibi Markov’un geliştirdiğı yöntemle çalışan n-gram günümüz teknolojisinde iletişimden, ulaşım, biyoenformatik alanından doğal dil işlemeye hemen hemen tüm alanlarda kullanılmaktadır.



İki durumlu Markov zinciri. Durumlar E ve A olarak adlandırılmıştır. Her bir sayı bir durumdan diğer duruma geçişte olasılıkları göstermektedir. Durumların geçişleri ok ile sembolize edilmiştir. Mesela, A durumundan E durumuna geçme ihtimali 0.4 iken aynı durumda kalma ihtimali 0.6'dır.

Şekil 3.3. İki durumlu Markov Modeli

N-gram, bir karakter katarının n adet karakter dilimidir. N-gram tabanlı sınıflandırma yöntemi, doküman içindeki n-gram karakterlerin kullanım sıklığına dayalıdır. Bu çalışmada, n-gram'ın farklı birkaç uzunluğu olarak 2-gram, 3-gram ve 4-gram'lar kullanılmıştır. N-gram'ların elde edilmesi bir örnekle açıklanacak olursa: Örnekte boşluk karakterini göstermek için “\_” altçizgi karakteri kullanılmıştır. Cümle “Yüksek Lisans” ise, bu cümlenin n-gram'ları;

2'nci gram'lar: "Yü", "ük", "ks", "se", "ek", "k\_", "\_L", "Li", "is", "sa", "an", "ns"

3'üncü gram'lar: "Yük", "üks", "kse", "sek", "ek\_", "k\_L", "\_Li", "Lis", "isa", "san", "ans"

4'üncü gram'lar: "Yüks", "ükse", "ksek", "sek\_", "sek\_L", "k\_Li", "\_Lis", "Lisa", "isan", "sans" şeklinde çıkarılır.

N-gram yöntemi, dokümanları sınıflandırmak için kullanılan güvenilebilir ve basit bir yöntemdir. Temel düşünce, bir doküman içindeki n-gram oluşumlarının tanımlanmasıdır. N-gram frekans yaklaşımı dilden bağımsız çalışır. Diğer bir ifade ile belli bir dil hakkında ayrıntılı bir dilbilgisine ya da bir sözlük yapısına gerek yoktur. Tüm harflerin ya da hecelerin istatistikleri kullanarak benzer sonuçlara ulaşılabilir. Bununla birlikte, bu düşünceyle beraber çok kısa paragraflardan oluşan dokümanlar, konu tabanlı kelime istatistiğinde oldukça yetersiz kalmaktadır. Neticede eşleme için önemli olabilecek N-gram'ların yeterli şekilde toplanabilmesi için daha uzun paragraflara gereksinim vardır. Sınıflandırma işlemi yazıdaki karakterlerin kullanım sıklığından yararlanılarak yapıldığı için konusu spor olan bir yazıda, bir kelimenin ilgili formları için ("futbol", "futbolcu", "futbolu", "futboldan") elde edilen n-gram'ların sıklığıyla sınıflandırma işlemi kolaylıkla yapılabilir. Doğal dillerin, kullanım sıklıkları diğerlerinden daha fazla olan değişmeyen bazı kelimeleri vardır. Bu varsayım bütün diller için geçerli bir varsayımdır. Bu düşüncenin ifadesinin en genel yollarından biri Zipf'in kanunlarıdır. Bu kanuna göre, bir dilde bazı kelimeler diğerlerinden daha çok kullanılır. Belli bir konuya özel olan kelimeler için bu doğru bir yaklaşımdır. Bu nedenle önemli frekans değerlerine sahip n-gram'ları diğerlerinden ayırmak için bir eşik değeri (threshold) kullanılır.

### 3.13. Naive Bayes Algoritması

Bayes teorimi, bir olayın kendisi ile ilişkili olayların gerçekleşme şartına bağlı olduğunu ortaya koyan istatistiksel bir teoridir. Dil işleme algoritmaları arasında verdiği sonuçlar sayesinde en popüler olan ve doğruluğa en yakın sonuç veren algoritmalarından biridir. Bayes teoreminde belli sınıflandırma özellikleri vardır. Bu sınıflandırma özelliklerinin en temel özelliği birbirinden tamamen bağımsız olmasıdır. Mesela bir meyvenin yuvarlak, kırmızı ve yaklaşık 10 cm yarıçapında olması onun elma olduğu anlamına gelebilir. Naive Bayes bu ilişkilendirmeyi yaparken bu özelliklerin birbirinden tamamen bağımlı olduğunu varsayar.

Yani meyvenin kırmızı renkli olması ne onun yarıçapı ile alakalıdır ne de yuvarlak olması ile alakalıdır.

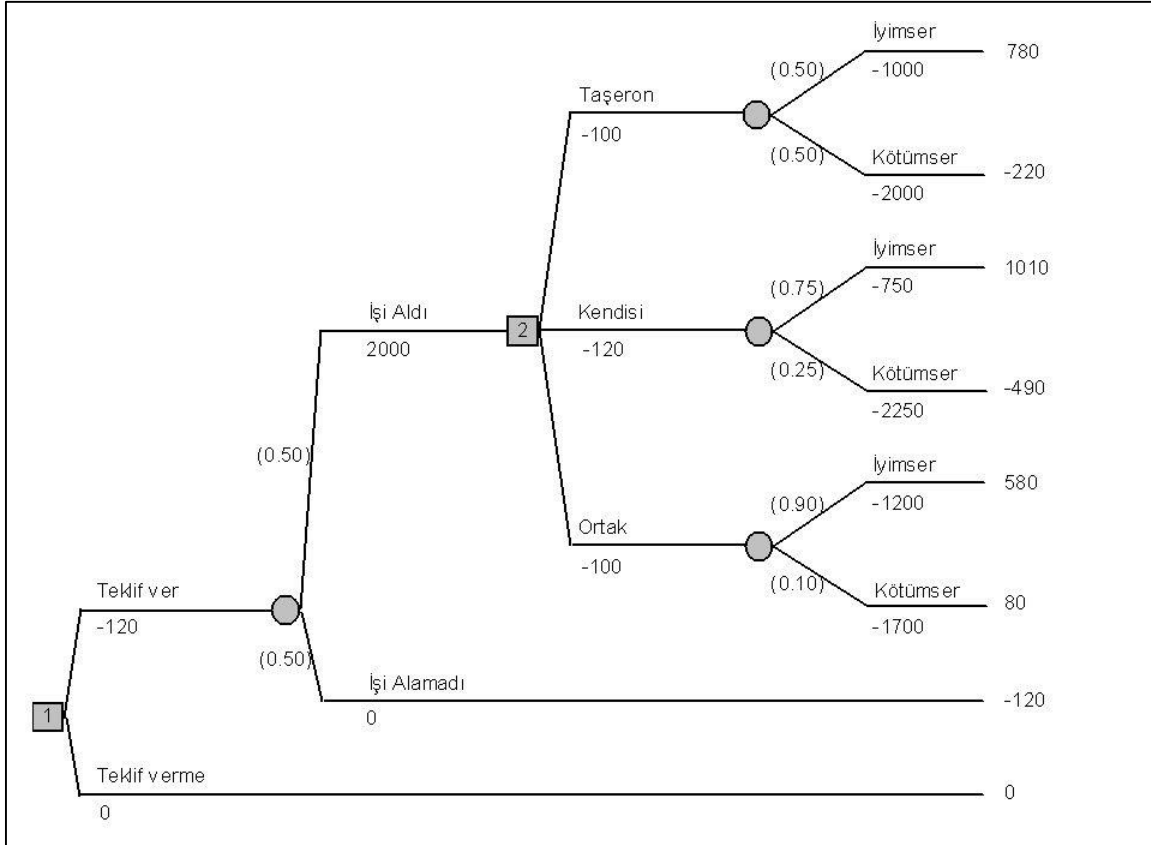
Karşılaştırmanın yapılabilmesi için bir sonraki kelimenin mutlaka bir öncekiyle bağlı şartı göz önüne alınarak yapılan bir karşılaştırma söz konusudur. Bu karşılaştırmaları yapabilmek için TextBlob [48] isimli Python kütüphanesi kullanılmıştır. Bu kütüphane NLTK [49] isimli doğal dil işleme kütüphanesini kullanmaktadır [50]. Bu karşılaştırma sayesinde en son kelime olarak etiket kabul edilmektedir. Etiketlerde ise cümlenin olumlu, olumsuz veya tarafsız olduğu ile ilgili 5 farklı etiket söz konusudur. Her bir tweet için ayrı ayrı sınıflandırma yapıp tüm etiketlere benzerlik oranı ondalık sayı olarak virgülden sonra 3 hanesi yuvarlanarak kullanıcıya gösterilmektedir.

Naive Bayes karşılaştırma çok küçük kütüphanelerde dahi doğruluk olarak son derece doğru sonuçlar verebilmektedir.

### **3.14. Karar Ağacı Algoritması**

Karar Ağacı algoritması oluşturulan sözlükteki her bir cümleyi ele alarak bir karar mekanizması oluşturmaktadır. Şekil 3.4'deki gibi ikili (doğru-yanlış Binary) şekilde her bir durum dallanmaktadır. En nihayetinde verilen girdiyi bize bu karar ağacı üzerinde yürüterek en doğru sonucu vermeyi amaçlamaktadır.

Bu algoritmanın kullanımı da TextBlob ve NLTK Kütüphanesi ile mümkün olmaktadır. TextBlob kütüphanesi NLTK sayesinde sahip olmuş olduğu DecisionTreeClassifier isimli sınıfı sayesinde kullanıcılara bu algoritma ile sınıflandırma yapabilme imkânı verebilmektedir [51]. Bu sınıfın bir özelliği de oluşturulan karar ağacı yapısını prettyprint isimli metodu ile verebilmesidir. Dikkat çeken en önemli özellik karar ağacı algoritmasının küçük kütüphanelerde bazen yanlış sonuçlar verse de kütüphanenin büyümesi halinde doğruluk değeri 1'e yani %100'e yaklaşmaktadır[52].



Şekil 3.4. Karar ağacı işlem mantığını gösteren örnek bir grafik

### 3.15. Programın Çalışma Algoritması

Twitter API'si (Şekil 3.5) üzerinden Twitter verileri üzerinde çalışma yapabilmek için öncelikle ücretsiz bir Twitter API anahtarı oluşturulması gerekmektedir.

```

consumer_key = 'ZaywEBd7uFXLJEOqgXmZCT18n'
consumer_secret = 'YUmIoXNHmfqh3mhiGEs4fcpFAqhMJxDTSd3HRYZRwREOyEBzAz'
access_token = '2271347449-xuNdTZfwVBXwfRm0klO6ozBcPwH9t8zKSrmCEox'
access_secret = 'O3AGtjI4k19aG0aEgXwCaMMXtzqKekmk1TDJzKLwWc1BU'

```

Şekil 3.5. Twitter API anahtarı

Uygulamada Twitter API ye ulaşabilmek için bir Python Kütüphanesi olan Tweepy kullanılmıştır. Tweepy aracılığıyla elde edilen verileri saklayabilmek için Postgre veri tabanı bağlantısıyla uygun modeller Şekil 3.6'daki gibi üretilmiştir.

```

admin.site.register(Ulke)
admin.site.register(Il)
admin.site.register(Ilce)
admin.site.register(MyTweet)
admin.site.register(HedefKisi)
admin.site.register(Hashtag)
admin.site.register(StopWords)

```

Şekil 3.6. Uygulama modelleri

Tweepy aracılığıyla *StreamListener* fonksiyonu kullanılarak canlı tweet akışı sağlanabilmektedir. Uygulamada hem canlı tweet akışı hem de belirli hashtagler ile ya da belirli kullanıcılara veya gruplara ait tweetler alınabilmektedir.

### 3.16. Tweetlerin Temizlenmesi ve Corpusa Eklenmesi

Modeller arasında bulunan MyTweet modelinde hem yalın hem de ilgili tweetin Şekil 3.7’de görüldüğü üzere temizlendikten sonraki hali de saklanmaktadır.

```

def clean(text):
    text = ' '.join(re.sub("(@[A-Za-z0-9İıÖöÜüĞğŞşÇç\ ]+)|([^\0-9A-Za-zİıÖöÜüĞğŞşÇç\ \t])|(\w+:\//\S+)",
        " ", text).split())
    try:
        patt = re.compile(u'(\u00002022)|([\u000025A0-\u000025FF])|([\u00002600-\u000027BF])|'
            u'([\u0001f300-\u0001f64F])|([\u0001f680-\u0001f6FF])')
    except re.error:
        patt = re.compile(
            u'(\u2022)|([\u25A0-\u25FF])|([\u2600-\u27BF])|([\uD83C][\uDF00-\uDFFF])|([\uD83D]'
            u'[\uDC00-\uDE4F])|([\uD83D][\uDE80-\uDEFF])')
    text = patt.sub('', text)
    text = re.sub(r'^RT[\s]+', '', text, flags=re.MULTILINE) # removes RT
    text = re.sub(r'https?:\//\.*[\r\n]*', '', text, flags=re.MULTILINE) # remove link
    text = re.sub('[^a-zA-Z0-9İıÖöÜüĞğŞşÇç\ ]+', '', text, flags=re.MULTILINE)
    text = " ".join(text.split())
    return text

```

Şekil 3.7. Tweetlerin temizlenmesi

Bu fonksiyonla tweetlerdeki gereksiz emojiler, RT etiketi, http linkleri, @ işareti silinmektedir. Sonuç olarak tweetlerden sadece alfa-numeric (alfabe ve rakamlardan oluşan katar) karakterler geriye kalmaktadır.

8	ID: 1101378485845745664	Türk Kadını Kucağındaki bebeğiyle birlikte acımasızca katleden PKK neden Siyasi Parti ve Feminist Dernekler üzerinden Kadın Haklarını savunuyor hiç düşündünüz mü	Türk Kadını Kucağındaki bebeğiyle birlikte acımasızca katleden #PKK neden #Siyasi Parti ve #Feminist Dernekler üzerinden Kadın Haklarını savunuyor hiç düşündünüz mü ? @RT_Erdogan @dbdevletbahceli @abdulhamitgul @ZehraZumurutS <a href="https://t.co/WIAdklMCAV">https://t.co/WIAdklMCAV</a>	01.03.2019	Ümraniye	Istanbul		
---	-------------------------	---	---	------------	----------	----------	---	---

Şekil 3.8. Yalın ve temiz tweet

Sisteme kaydedilen her tweet sistemde "Raw Tweet" sekmesi altından erişilebilmektedir. Şekil 3.8'de görüldüğü gibi her tweet tweetin ID'si, o tweeti atan kullanıcının ekran ismi (screen\_name) ve eğer mümkünse yani kullanıcı Twitter ile kendi konum bilgisini paylaşmış ise konum bilgisi saklanmaktadır. Sisteme kayıtlı bu tweetler aslında tek başına bir anlam ifade etmemektedir. Sisteme kayıtlı tweetin analiz ve eğitim (sistem eğitilmesi) işleminde kullanılabilmesi için mutlaka Corpus adını verdiğimiz sisteme eklenmesi gerekmektedir. Bu işlem sistemde "Raw Tweet" sekmesi altında Corpus Ekle butonu ile manuel olarak yapılmaktadır.

### 3.17. Test ve Train Dataset Oluşturulması

Corpus'a kayıtlı 14048 tweetin %70'i test amaçlı %30'u da eğitime amacıyla kullanılmaktadır. Bu işlemin gerçekleşebilmesi için "test\_and\_training\_data\_set" fonksiyonu kullanılmaktadır. Öncelikle Corpus'a kayıtlı tüm tweetleri zaman damgasına göre en son oluşturulan tweet en başta olacak şekilde sorgu işlemi yapılarak veri tabanı kayıtlı MyTweet modelinde her bir tweet için type sütununa test mi yoksa eğitime amacıyla kullanılacağı bilgisi girilmektedir.





Şekil 3.9. Programın iş akış şeması

### 3.18. Programın İş Akışı

Şekil 3.9'da görüldüğü üzere bir kişinin tweetlerini analiz etmek için ilk başta kullanıcı adı ve şifreyle giriş yapılmalı. İkinci aşamada analizi yapılacak olan kişi takip listesine eklenmeli. Üçüncü adımda API üzerinden listedeki kişilerin verileri çekilir ve veri tabanına kaydedilir. Dördüncü adımdaysa kaydedilen veriler 3 algoritmayla analiz edilir ve elde

edilen sonuçlar veri tabanına yazılır. Beşinci adımda sonuçlar ister tablo halinde isterse de harita üzerinde gösterilir.





## 4. ARAŞTIRMA BULGULARI

Uygulamanın anlamlı sonuçlar çıkarabilmesi eğitim kümesinin'in oldukça büyük olması gerekmektedir. Analiz işlemleri NLTK ve SklearnClassifier Kütüphanesi kullanılarak 3 algoritma kullanılmış ve sistem eğitilmiştir:

```
NB_classifier = nltk.classify.NaiveBayesClassifier.train(trainList)
DT_classifier = nltk.classify.DecisionTreeClassifier.train(trainList)
SC_classifier = SklearnClassifier(BernoulliNB()).train(trainList)
```

Sistemin doğru sonuçlar verebilmesi corpusta yer alan eğitme datasetin bigram\_vectorizer kullanılarak vektörel boyut kazandırılmasına bağlıdır.

```
bigram_vectorizer = CountVectorizer(ngram_range=(1, 2), token_pattern=r'\b\w+\b',
min_df=1)
```

Vektörizasyon da denilen işlemde algoritmaların doğru çalışabilmesi için text içerikli katarlar sadece sayılarla ifade edilmektedir. Sistem eğitme içerisindeki kelimelerden kabul edilebilir olanları yani stop words (çok sık kullanılan ve veya, and, or, the gibi edatlar) elendikten sonra her bir kelimeye bir id atamakta ve ilgili cümle (kelime grubu) sadece bu sayılardan oluşmaktadır.

Sistemde Naive Bayes, DecisionTree ve SklearnClassifier (N-Gram) kullanılmaktadır:

### 4.1. Naive Bayes Algoritması ve nltk.NaiveBayesClassifier

Classifier, adından da anlaşılacağı gibi sınıflandırıcıdır. Twitter API den elde edilen yalın tweetler Corputa kaydedilirken mutlaka onlara 1-5 arası bir etiket (etiket) verilmektedir. Bunlar; 1- Terör örgütü Karşıtı, 2- Terör karşıtı ama liberal, 3- Liberal, 4- Terör yanlısı ve liberal ve 5-Terör örgütü Sempatizanı şeklindedir. Kullanıcı sistemin doğru çalışabilmesi için bu bölümde (Şekil 4.1) oldukça seçici ve dikkatli olunmalıdır. Varsayılan olarak Corputa eklenen her tweet 3-Liberal, görüş bildirmiyor etiketi ile eklenmektedir.

Tweet-Corpusa Ekle

Vazgeç

---

**Kullanıcı** ID: 1101378626291945472

ScreenName:fatma268054

Location:Sakarya, Türkiye

**Clean** DEMOKRASI PKK'NIN DON LASTIĞI OLMAKMIŞ CHP'YE GÖRE HAKLISINIZ ÇÜNKÜ ARTIK KOL KOLASINIZ  
ZilletIttifakınınKandilAşkı

**RawTweet** DEMOKRASI PKK'NIN DON LASTIĞI OLMAKMIŞ CHP'YE GÖRE HAKLISINIZ ÇÜNKÜ ARTIK KOL KOLASINIZ  
#ZilletIttifakınınKandilAşkı <https://t.co/mRzeDnJDGB>

**Tweet-Zamanı** 01.03.2019

**Değerlendirme** 1-  2-  3-  4-  5-

1- PKK karşıtı;2-PKK karşıtı ancak liberal; 3--Liberal , Görüş Bildirmiyor; 4--PKK Sempatizanı;5-PKK üyesi

Corpusa Ekle

Şekil 4.1. Corpusa ekle

NLTK Naive BAYES Classifier algoritması temel olarak Naive Bayes teoremini kullanarak birbirinden bağımsız değişkenler üzerinde herhangi bir etiketin o elemana ait olup olmadığını araştırır. Uygulamada 5 etiket olduğu için bu classifier her bir etiket için ayrı ayrı çağrılır ve ilgili tweetin hangi etiket ile ne oranda ilişkili olduğu bulunur.

```
nb_analiz = models.FloatField(default=0.0)
nb_analiz1 = models.FloatField(default=0.0)
nb_analiz2 = models.FloatField(default=0.0)
nb_analiz3 = models.FloatField(default=0.0)
nb_analiz4 = models.FloatField(default=0.0)
nb_analiz5 = models.FloatField(default=0.0)
```

Yukarıda da görüleceği üzere MyTweet modelinde her bir etiket için ayrı bir alan mevcuttur. nb\_analiz alanı ise sonuç olarak o tweetin en çok hangi etikete ait olduğunu belirtir.

```
nb_analiz = Max(nb_analiz1, nb_analiz2, nb_analiz3, nb_analiz4, nb_analiz5)
```

## 4.2. Karar Ağacı

Karar Ağacı algoritması olarak Türkçeye çevrilen bu algorithmada öncelikle tüm olasılıklar değerlendirilerek dallanmalar yapılarak bir ağaç modeli ortaya konur. Sonuç olarak herhangi bir tweetin neye uygun olduğu konusunda ağaç üzerindeki bu dallanmalar takip edilerek kesin bir sonuç ortaya konur. Decision Tree algoritması kullanılarak oluşturulan bu cassifier'da tüm etiketler ile ilgili olasılık söz konusu değildir. Adından da anlaşılacağı üzere Karar Ağacı denilen bu sınıflandırıcı da ancak ve ancak tek bir sonuç üretilmektedir. Yani sınıflandırıcı da herhangi bir tweetin %100 olarak hangi etiket altında olduğu bulunur. Dolayısıyla geri kalan 4 etiket %0 olasılığa sahip olacaktır.

## 4.3. Sklearn Classifier (BernoulliNB)

Scikit-learn (sklearn) bir Python sınıflandırıcı modülüdür. İçerisinde Naive Bayes sınıflandırıcı temelli ancak optimize edilmiş MultinomialNB, BernoulliNB ve SGDClassifier isimli sınıflandırıcılar vardır. Çalışmada Naive Bayes teoremi tabanlı ve çok iyi şekilde optimize edilmiş olan BernoulliNB isimli classifier kullanılmıştır. Bunun sebebi diğer sınıflandırıcılara nazaran doğruluk oranının %100'e yakın olmasıdır. BernoulliNB bir N-gram makine eğitimi ve sınıflandırma algoritmasıdır.

## 4.4. Uygulamanın Ekran Çıktıları

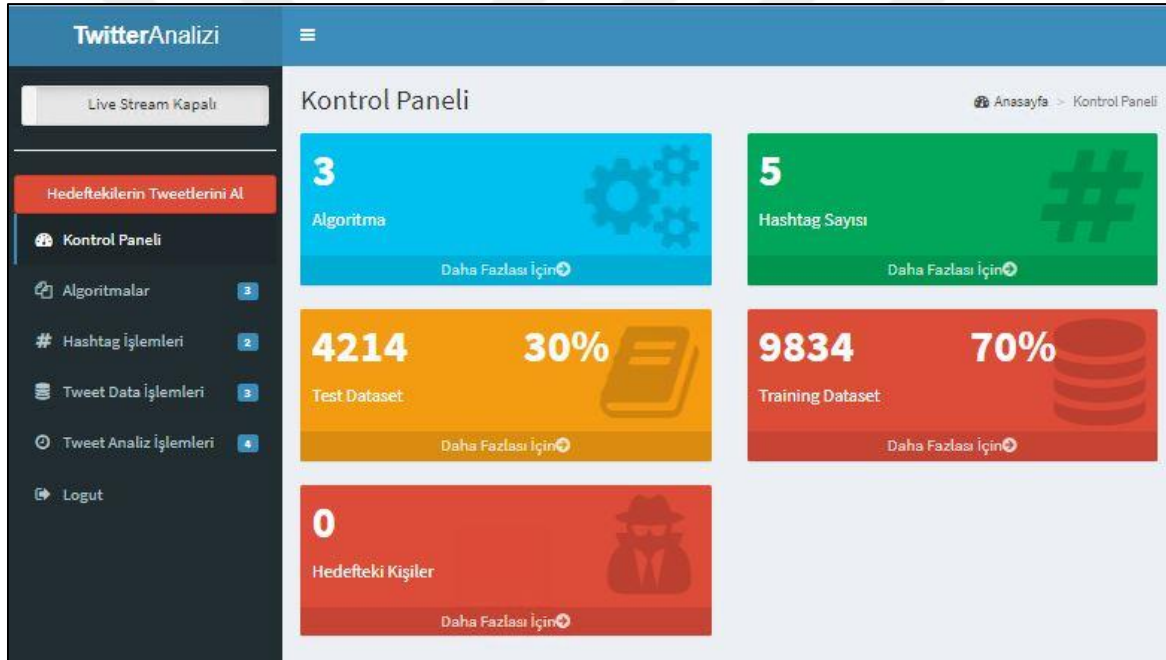
The image shows a login form with the following elements:

- Title: **Sisteme Giriş**
- Input field for **Kullanıcı Adı** (Username)
- Input field for **Şifre** (Password)
- Button: **Giriş Yap** (Login)

Şekil 4.2. Giriş sayfası

Şekil 4.2'deki gibi bu ekranda kullanıcı adı ve şifre ile sisteme giriş yapılmaktadır. Kullanıcı adı ve şifre Django altyapısı vasıtasıyla daha önceden belirlenmektedir. Yetkilendirme sınıfları sıfırdan yazılmamıştır. Bütün yetkilendirme ve sayfalara erişim sınıfları Django kütüphanesinin standart metotlarını kullanmaktadır.

Sisteme giriş yaptıktan sonra Şekil 4.3'de görüldüğü gibi yönetim paneli görülmektedir. Yönetim panelinde birden fazla menü ve gösterge bulunmaktadır. Sol taraftaki menülerde erişilecek kısımların listesi mevcuttur. Bu menülerden algoritmalara veri tabanlarına, takip edilen kişilere ve test veri tabanına erişilebilir. Ana ekrandaysa kaç tane Algoritmanın çalıştığı kaç adet tweetin toplandığı ve kaç tane tweet işlendiği görülebilmektedir.



Şekil 4.3. Ana yönetim paneli

Naive Bayes Algoritması

Anasayfa > Algoritmalar > Naive Bayes Sınıflandırıcı

## Naive Bayes Sınıflandırıcı

### Naive Bayes sınıflandırıcı

Kaynak = Vikipedi, özgür ansiklopedi



Bayes Teoremi

**Naive Bayes Sınıflandırıcı** adını İngiliz matematikçi Thomas Bayes'ten (yak. 1701 - 7 Nisan 1761) alır. Naive Bayes Sınıflandırıcı örüntü tanıma problemine ilk bakışta oldukça kısıtlayıcı görülen bir önerme ile kullanılabilen olasılıkçı bir yaklaşımdır. Bu önerme örüntü tanıma da kullanılacak her bir tanımlayıcı öznelik ya da parametrenin istatistik açıdan bağımsız olması gerekliliğidir. Her ne kadar bu önerme Naive Bayes sınıflandırıcının kullanım alanını kısıtlasa da, genelde istatistik bağımsızlık koşulu esnetilerek kullanıldığına da daha karmaşık yapay sinir ağları gibi metotlarla karşılaştırılabilir sonuçlar vermektedir. Bir Naive Bayes sınıflandırıcı, her özneliğin birbirinden kosulsal bağımsız olduğu ve öğrenilmek istenen kavramın tüm bu özneliklere kosulsal bağılı olduğu bir Bayes ağı olarak

Şekil 4.4. Naive bayes algoritması

Algoritma menüsü, bu menüde kullanılan birkaç algortmadan en fazla cevap alınan algortmalarının kısa özeti ve tanıtımı yapılmıştır.

Decision Tree(Karar Ağacı) Algoritması

Anasayfa > Algoritmalar > Decision Tree Algoritması

## Decision Tree Algoritması

### Karar Ağacı Öğrenmesi (decision tree learning)

Şadi Evren ŞEKER

Karar ağacı öğrenmesi (decision tree learning) yöntemi, makine öğrenmesi (machine learning) konularından birisidir. Literatürde karar ağacı öğrenmesinin alt yöntemleri olarak kabul edilebilecek sınıflandırma ağacı (classification tree) veya ilkelleştirme ağacı (regression tree, tahmin ağacı) gibi uygulamaları vardır.

Karar ağacı öğrenmesinde, bir ağaç yapısı oluşturularak ağacın yaprakları seviyesinde sınıf etiketleri ve bu yapraklara giden ve başlangıçtan çıkan kollar ile de özellikler üzerindeki işlemler ifade edilmektedir.

Bu yazı karar ağacı olarak geçen (decision tree) konusunu değil, veri madenciliği (data mining) ve makine öğrenmesi (machine learning) konularında geçen, karar ağacı öğrenmesi ile ilgilidir.

Karar ağacı öğrenmesi sırasında, öğrenilen bilgi bir ağaç üzerinde modellenir. Bu ağacın bütün iç düğümleri (interior nodes) birer girdiyi ifade eder. Örneğin aşağıda örnek bir öğrenilmiş ağaç gösterilmiştir:

Yukarıdaki şekil, kalp krizi geçiren kişilerin çeşitli özellikleri ele alınarak veriler üzerinden öğrenilmesi durumunda hayali olarak elde edilecek sonuçlardır.

Şekil 4.5. Karar ağacı algoritması



N-gram bir sınıflayıcı değildir, istatistiksel bir dil modelleme aracıdır. Bu modellemede temel birimler üzerine kurgulanmış bir API söz konusudur. Bu temel birimler kelime, kelime grubu ya da harf olabilir. Bu modelleme bir çok sınıflayıcı algortmada kullanılan bir yöntem olmuştur. Sınıflayıcılar bu modellemeyi kulnır ya da kullanmaz. N-gram "n"(n bir doğal sayı 2-3-4-5-6...) olabilir. Çalışmamız sırasında Nltk kütüphanesinden NaiveBayesClassifier,DecisionTreeClassifier,SklearnClassifier(BernoulliNB()) ve sklearn kütüphanesinden sklearn.tree.DecisionTreeClassifier kullanılmıştır.

```

NB_classifier = nltk.classify.NaiveBayesClassifier.train(trainList)
DT_classifier = nltk.classify.DecisionTreeClassifier.train(trainList)
SC_classifier = SklearnClassifier(BernoulliNB()).train(trainList)

```

```

features = CountVecorizer(ngram_range=(1, 2), token_pattern=r'\b\w+\b', min_df=1)
features = bigram_vectorizer.fit_transform([[xx[0] for xx in corpus]].toarray())

```

sklearn.tree.DecisionTreeClassifier algortması sırasında Bi-Gram Vektörler kullanılmıştır.

Yazan: Şadi Evren ŞEKER

Verilen bir dizilimdeki (sequence) tekrar oranını bulmaya yarayan yöntemdir. İsmi n ve gram kelimelerinin birleşiminden oluşmaktadır. Buradaki n, tekrarı kontrol edildiği değerdir. Gram ise bu tekrarı dizilim içerisindeki ağırlığını ifade etmek için kullanılmıştır.

Örneğin bir dizi (string) içerisindeki n-gram değerini bulmak isteyelim ve buradaki n değeri 2 olsun (n = 2)

Yukardaki yazının 2-gram değerleri aşağıda verilmiştir:

Şekil 4.6. N-gram

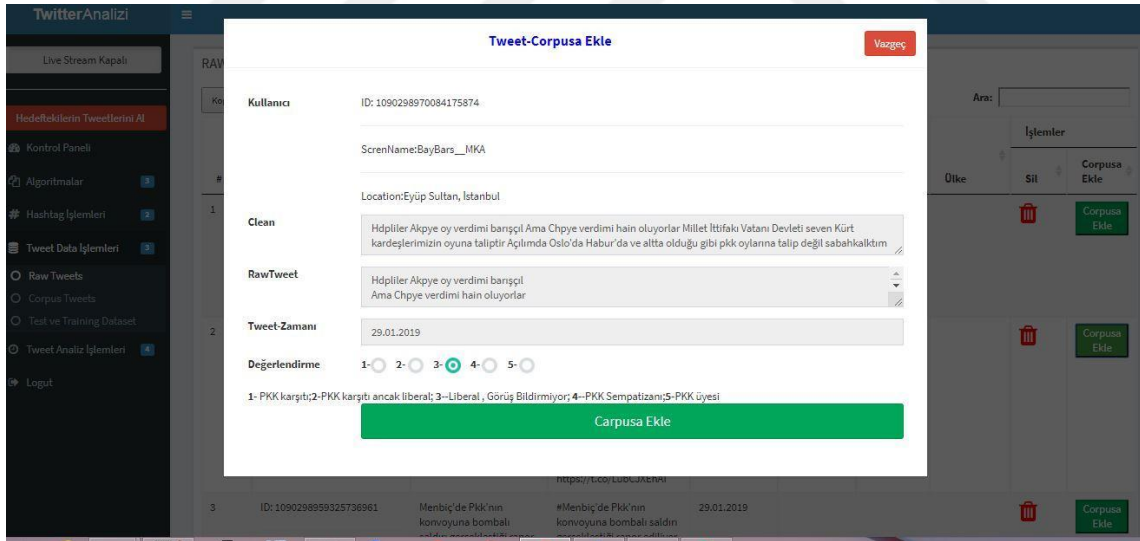
Şekil 4.4, Şekil 4.5 ve Şekil 4.6'da görüldüğü üzere bu alt menüdeyse Karar Ağacı algortması, Naive Bayes Algortması ve N-gram hakkında kısa ve özet bilgi verilmiştir. Ayrıca bu algortmanın nasıl çalıştığına dair kısa bilgi sunulmuştur.

RAW TWEETLER									
Ara: <input type="text"/>									
#	Kullanıcı	CleanTweet	RawTweet	Zaman	İlçe	il	Ülke	İşlemler	
								Sil	Corpusa Ekle
1	ID: 1090299056142704641 ScreenName:BasGazetesi Location:	Pentagon Suriye Demokratik Güçleri IŞİD'in elindeki son toprakları birkaç haftada geri alır	Pentagon: Suriye Demokratik Güçleri IŞİD'in elindeki son toprakları birkaç haftada geri alır <a href="https://t.co/lhmrzfuupC">https://t.co/lhmrzfuupC</a>	29.01.2019					<a href="#">Corpusa Ekle</a>
2	ID: 1090299022890422272 ScreenName:Mr_garko Location:Kano	6 bedroom duplex with pent at apo resettlement N52m	6 bedroom duplex with pent at apo resettlement N52m <a href="https://t.co/CjeBvmsD0D">https://t.co/CjeBvmsD0D</a>	29.01.2019					<a href="#">Corpusa Ekle</a>
3	ID: 1090299018633203719	No I am an MDC voter one of the people who gave	@herbert_gomba @nicolehonda No I am an	29.01.2019					<a href="#">Corpusa Ekle</a>

Şekil 4.7. Yalın tweet

Yalın tweet ekranından takip edilen kişilerin tweetleri web AI aracılığıyla çektikten sonra veri tabanına eklenir, veri tabanına ekledikten sonra ise bu tweetler yalın tweet olarak algılanmaktadır. Yalın tweet daha puanlanmamış tweet demektir. Bu ekran kullanılarak insani duygularla tweetlere puan verilir ve değerlendirilir. Daha sonra bu yalın tweetler ve verilmiş puanlar analiz sürecinde kullanılmıştır.

Tweetleri puanlama ekranı, Şekil 4.8'deki gibi bu ekranda yalın tweetleri seçerek onlara birden beşe kadar puan verilmektedir. Bir (1) tamamen terör örgütüne karşı olan bir tweet demektir. Beş (5) ise tamamen terör örgütüne yakın ve neredeyse terör örgütü üyesi konumunda olan kişinin tweeti demektir. Ayrıca bu tweetleri puanladıktan sonrada verilen puanları aynı ekrandan güncelleme imkânı sunulmaktadır. Bu ekran özel önem arz etmektedir zira ki bu ekranda puanlanan tüm tweetler ileride sitemin temelinin ve algoritmaların girdilerini oluşturacağı için bir uzman kişi veya kişiler tarafından puanlanmalıdırlar. Eğer bu aşamada bir uzman insan tarafından bu puanlama yapılmazsa ileride algoritma hedeflenen performansı ve çıktıyı veremeyecektir.



Şekil 4.8. Yalın tweet

CORPUS TWEETLER										
Ara: <input type="text"/>										
#	ID	CleanTweet	RawTweet	Zaman	Puan	İlçe	İl	Ülke	Sit	Güncelle
1	ID:1090299056142704641 ScreenName:BasGazetesi Location:	Pentagon Suriye Demokratik Güçleri IŞID in elindeki son toprakları birkaç haftada geri alır	Pentagon: Suriye Demokratik Güçleri IŞID'in elindeki son toprakları birkaç hafta geri alır https://t.co/lhmrfzuupC	29.01.2019	3					
2	ID:1090299018633203719 ScreenName:HelenMutsvairo Location:Harare, Zimbabwe	No I am an MDC voter one of the people who gave you the mandate You need to be responsible why would a Democratic Party ask for destruction of people s livelihood withought a court order Musaita chihunhu chechi	@herbert_gomba @nicolehondo No I am an MDC voter one of the people who gave you the mandate . You need to be responsible ,why would a Democratic Party ask for destruction of people's livelihood withought a court order! Musaita chihunhu chechi zanu apo	29.01.2019	3					

Şekil 4.9. Corpus-tweet

Şekil 4.9'daki gibi yalnız tweetler puanlandıktan sonra Corpus'a eklenir. Corput'a puanlanmış olan tweetler vardır. Üstteki şekilde görüldüğü gibi Corputaki tweetleri görebilir ve gerektiğinde onları güncelleyebilir veya silebilir. Tweetler Corputa eklenme esnasında daha önce belirtilmiş temizleme aşamasından geçirilir ve gereksiz olan karakterler, emojiler, özel karakterler ve Hashtag'lar silinir. Bu temizlenmiş tweet ise veri tabanında ayrı bir alana yazılır.

TwitterAnalizi

Live Stream Kapalı

Hedeflekilerin Tweetlerini Al

Kontrol Paneli

Algoritmalar

Hashtag İşlemleri

Tweet Data İşlemleri

Raw Tweets

Corpus Tweets

Test ve Training Dataset

Tweet Analiz İşlemleri

Logout

**Tweet-Corpusa Güncelle** Vazgeç

**Kullanıcı** ID: 1090299018633203719

ScreenName:HelenMutsvairo

Location:Harare, Zimbabwe

**Clean**  
No I am an MDC voter one of the people who gave you the mandate You need to be responsible why would a Democratic Party ask for destruction of people s livelihood withought a court order Musaita chihunhu chechi

**RawTweet**  
@herbert\_gomba @nicolehondo No I am an MDC voter one of the people who gave you the mandate . You need to be responsible ,why would a Democratic Party ask for destruction of people's livelihood withought a court

**Tweet-Zamanı**  
29.01.2019

**Değerlendirme** 1-0 2-0 3-0 4-0 5-0

1- PKK karşıtı;2-PKK karşıtı ancak liberal;3-Liberal ; Görüş Bildirmiyor;4-PKK Sempatizanı;5-PKK üyesi

Corpusa Ekle

Ara:

İşlemler

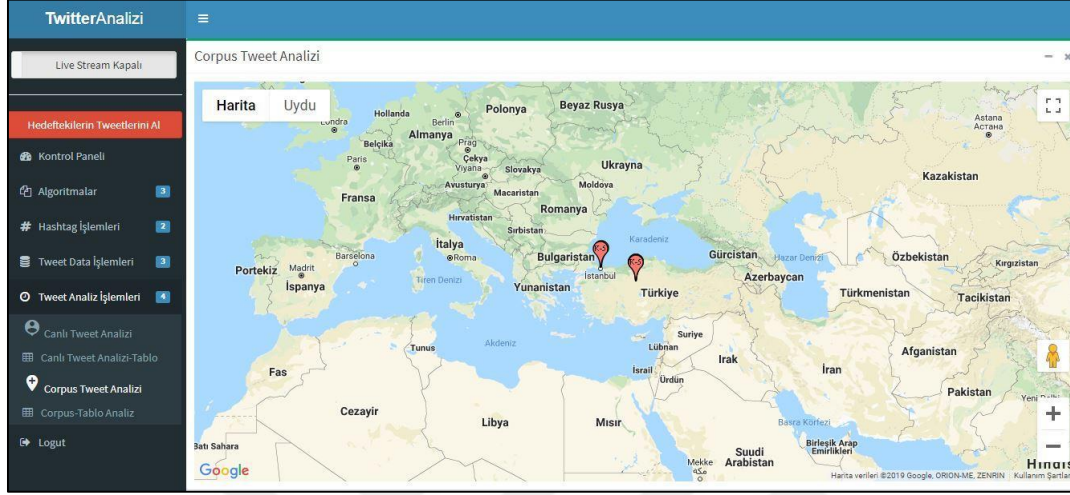
Ülke

Sit

Güncelle

Şekil 4.10. Corpus-güncelle

Şekil 4.10'da görüldüğü gibi bir kayıt eklendikten sonra aynı sayfada tekrardan güncellenebilir. Bu ise sitemdeki kullanıcıların olası yanlış değerlendirmelerini düzeltme imkânı sunmaktadır.



Şekil 4.11. Corpus-analiz

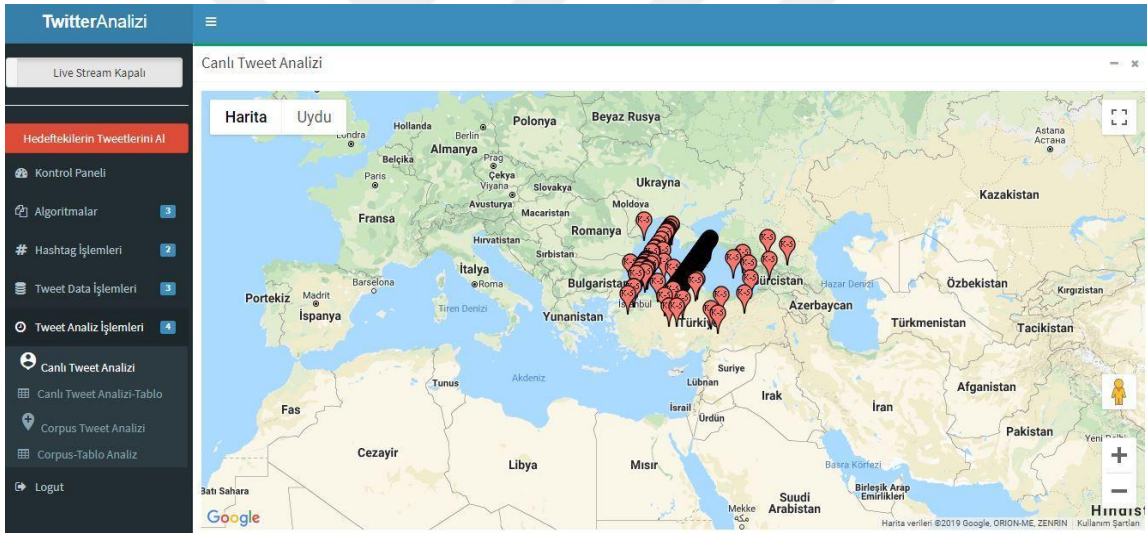
Corpus'taki tweetler algoritmalarla analiz edilebilir. Bu analiz ise algoritmaların doğru çalışıp çalışmadığına dair bir fikir sunmaktadır. Şekil 4.11'de görüldüğü gibi Corpus'taki tweetler analiz edildikten sonra haritanın üzerinde atılan tweetlerin konumu ve kişilerin isimleri görülmektedir. Bir kişinin hangi kategoriye ait olduğuyorsa toplamda 3 algoritmanın çıktılarının ortalamasını alınarak karar verilmektedir. Bu çıkarım yapıldıktan sonra kişinin konumu son attığı tweetten alınarak kişinin son konumu Google Maps aracılığıyla gösterilmektedir.

CORPUS TWEETLER										
#	ID	CleanTweet	RawTweet	Zaman	Puan	İlçe	İl	Ülke	İşlemler	
1	ID: 1090299056142704641 ScreenName:BasGazetesi Location:	Peritagon Suriye Demokratik Güçleri IŞİD'in elindeki son toprakları birkaç haftada geri alır	Peritagon: Suriye Demokratik Güçleri IŞİD'in elindeki son toprakları birkaç haftada geri alır https://t.co/InmrtzuupC	29.01.2019	3					
2	ID: 1090299018633203719 ScreenName:HelenMutsavairo Location:Harare, Zimbabve	No I am an MDC voter one of the people who gave you the mandate You need to be responsible why would a Democratic Party ask for destruction of people's livelihood without a court order Musait chihunhu chechi	@herbert_gomba @nicolehondo No I am an MDC voter one of the people who gave you the mandate. You need to be responsible why would a Democratic Party ask for destruction of people's livelihood without a court order Musait chihunhu chechi zanu apo	29.01.2019	3					

Şekil 4.12. Corpus-analiz-tablo

Aynı Corpus ve aynı tweetler bu sefer harita üzerinde değil de tablo şeklinde (Şekil 4.12) yapılırsa oranlar ve çıkarım sonuçları rakamsal olarak gösterilebilir.

Canlı tweet analizinde bugüne kadar veri tabanında olan bütün tweetler ve hatta web Api üzerinden çekilen tweetler anlık olarak sınıflandırılmaktadır. Eğer bir kişinin veya takip edilen kişinin tweetleri analiz edilmek isteniyorsa ilk önce o kişi takip listesine eklemeli ve o kişinin geriye dönük 200 tweeti ilk önce veri tabanına yazılmalıdır. Veri tabanına yazdıktan sonra canlı tweet analizi ekranından bu tweetler analiz edilebilir ve kişinin hangi örgüte yakın olduğu analiz edilebilir. Şekil 4.13’de görüldüğü gibi takip edilen kişiler veya Hashtag vasıtasıyla uygulamadaki filtreye takılan kullanıcıların tweetleri alınıp, işlenir. Bu aşama görece olarak uzun süre alır zira her 3 algoritmayı çalıştırarak yüksek miktarlarda işlemciye yük bindirir.



Şekil 4.13. Canlı-analiz-tablo



TwitterAnaliZi

Live Stream Kapalı

Tüm Karşılaştırmalar NaiveBayes DecisionTree SklearnClassifier Analiz

K1 =PKK karşıtı  
K2 =PKK karşıtı ancak liberal  
K3 =Liberal, Görüş Bildirmiyor  
K4 =PKK Sempatizani  
K5 =

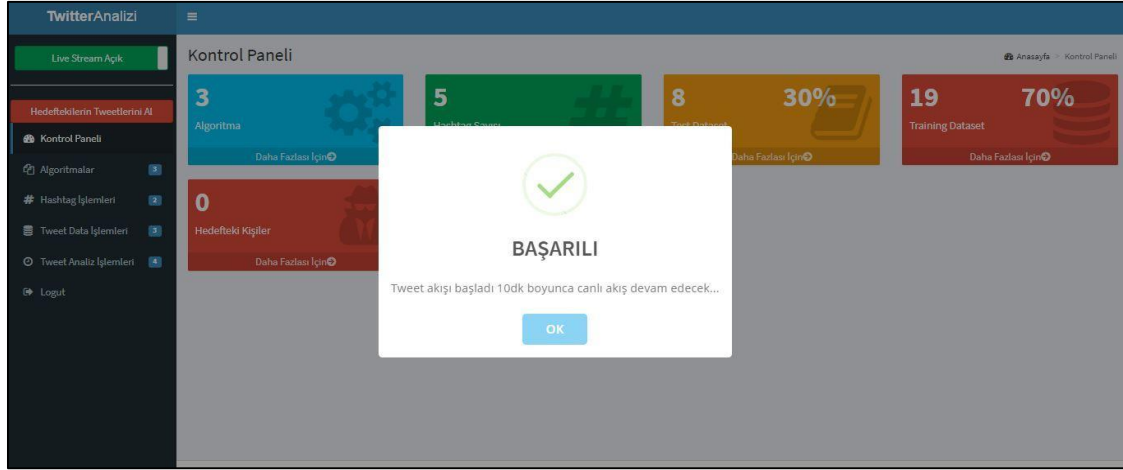
Kopyala Excel CSV PDF

Ara:

#	User	CleanTweet	Naive Bayes Analiz DO:0.0%					Decision Tree Analiz DO:0.0%					SklearnClassifier Analiz DO:100.0%				
			K1	K2	K3	K4	K5	K1	K2	K3	K4	K5	K1	K2	K3	K4	K5
1	Mr_garko	6 bedroom duplex with pent at apo resettlement NS2m	0.13	0	0.394	0.083	0.589	0	0	0	0	0	0	0	0.5	0.0	0.5
2	sawutul_ummah3	Menbiç'de Pkk'nin konvoyuna bombalı saldırı gerçekleştiği rapor ediliyor	0.13	0	0.394	0.083	0.589	0	0	0	0	0	0	0.5	0.0	0.5	
3	hasanyalcin47	1 Cümledeki anlatım bozukluğu yılların gazeteci yazarna yakışmıyor 2 Karşı çıkan HDP'ilerin hiçbirini jstci değil tam	0.13	0	0.394	0.083	0.589	0	0	1	0	0	0.0	0	0.5	0.0	0.5

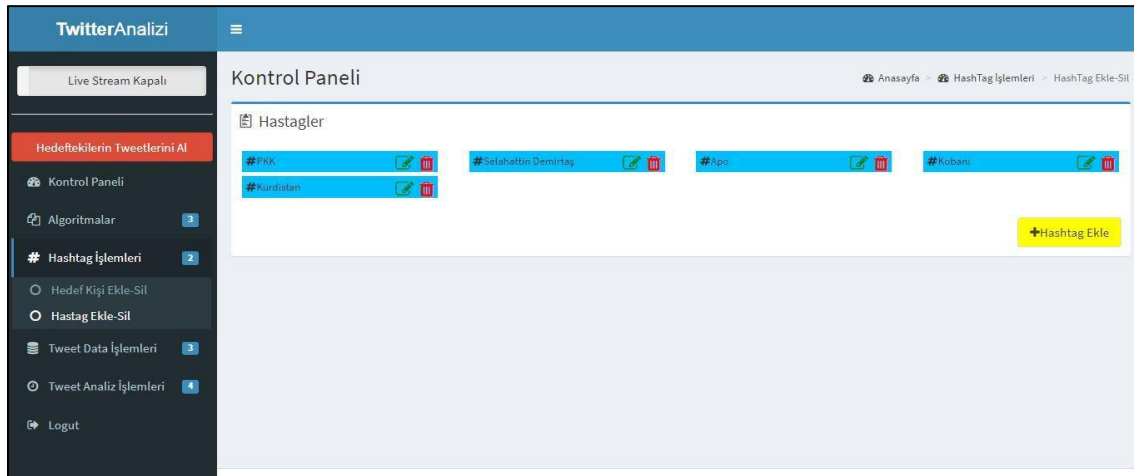
Şekil 4.14. Canlı-analiz-tablo

Aynı analiz işlemi harita üzerinden değil de tekrardan bir tablo olarak görmek istenirse Şekil 4.14'deki gibi Canlı Twitter analizi tablo olarak yapılması gerekecektir. Tablo şeklindeki yapılan analizler detaylı bilgi sunmaktadır. Hangi tweetinin ne kadar puan aldığı ve hangi örgüte yakın olup olmadığı buradan görülebilir. Bu ekranda 3 sekme mevcuttur. Her bir algoritma için nihai çıktı tweet temelli gösterilmektedir. Şekil 4.14'de görüldüğü gibi bir kişinin veya tüm kullanıcıların yazdıkları teker teker puanlanır. Eğer önceki adımlarda Corpus doğru oluşturulduysa bu ekranda gerçeğe yakın çıktılar görülebilir. Bir tweet aynı zamanda birden fazla kategoriye girme şansı vardır. Ama bu mensubiyet 0 ile 1 arasında (örneğin 0.589) bir rakamla ifade edilir. Alınan en yüksek rakam ise o tweet için son karar olarak temel alınır.



Şekil 4.15. Canlı-tweet-akışı

Uygulamada daha önce değinildiği gibi birden fazla yöntemle canlı olarak tweetleri çekebilir. Bu yöntemlerden biri Hashtag yöntemidir. Bu çalışmada bir örgüt jargonuna yakın kelimeler sisteme girilir. Oluşan veri kümesi o örgütün tweet atma biçimi ve anahtar kelimelerini temsil eder. Bu kelime kümesi oluştuktan sonra canlı tweet çekme işlemini Şekil 4.15’de görüldüğü gibi ekranın sol üst köşesinden başlatılır. Bu işlem başlatıldıktan sonra sistem 10 dakika boyunca daha önce belirlenen Hashtag’leri içeren tweetleri anlık olarak Twitter sunucusundan çeker ve veri tabanına kaydeder. Bu tweetler doğal olarak ilk aşamada yalnız tweet olarak muamele görür. Bu tweetlerin arasına çoğu zaman alakasız tweetlerde karışabilir. Bu sorunu gidermek içinse canlı tweet çekme işlemi sona erdikten sonra gereksiz tweetler yalnız tweet ekranından silinir.



Şekil 4.16. Hashtag ekleme ekranı

Sistemde Şekil 4.16'da görüldüğü gibi tweetleri Twitter'dan çekme şansı vardır. Birincisi Hashtag tabanlı ve ikincisi de kişi tabanlıdır. Örnek olarak eğer #rojhalat kelimesini sisteme kaydedilirse, Twitter bu kelimeyi içeren son 200 tweeti verecektir. Bazen aynı Hashtag altında karşıt görüşler yazılır. Bu sebepten yalın tweet ayıklama işlemi çok dikkatli yapılmalıdır. Aksi takdirde çıkarım sonuçları tamamen farklı ve anlamsız olacaktır. Bunun alternatif çözümü ise bir örgüte mensup olduğundan emin olunan kişiyi takibe almaktır. Bu kişi genellikle o örgütün propaganda amaçlı tweetlerini paylaşır. Yani o kişinin attığı tweetlerin büyük çoğunluğu örgütün jargonuna uymaktadır. Bu ise Yalın tweet ayıklama esnasında büyük kolaylık sağlamaktadır. Tweeti atan kişinin adını gördükten sonra ve tweetin birkaç satırını okuduktan sonra hangi örgüte ait olup olmadığı ve ne kadar puan verilmesi gerektiği rahatlıkla anlaşılabilir.

#### **4.5. Canlı Tweet Analizi**

Canlı tweet analizi live stream ile elde edilen en son 200 tweet üzerinde gerçekleştirilmektedir. Uygulamada canlı tweet analizi hem harita üzerinde görüntülenebilmekte hem de tablo şeklinde görüntülenebilmektedir. Analizler tek bir tabloda yan yana doğruluk oranları da belirtilerek incelenebilmekte veya yan sekmede şekil 4.18 – 4.20'deki gibi her bir sınıflandırıcı için sonuçlar tek tek görüntülenebilmektedir. Programın çalışma algoritmaları yukarıda detaylı olarak anlatılmıştır. Şekil 4.20 ve Şekil 4.21'de ise canlı tweet analizinin harita üzerinde ve tablo halinde gösteriminin ekran görüntüleri verilmiştir.



TwitterAnalizi

Live Stream Kapalı

Hedeflekilerin Tweetlerini Al

Kontrol Paneli

Algoritmalar 3

Hashtag İşlemleri 2

Tweet Data İşlemleri 3

Tweet Analiz İşlemleri 4

Canlı Tweet Analizi

Canlı Tweet Analiz-Tablo

Corpus Tweet Analizi

Corpus-Tablo Analiz

Logout

Tüm Karşılaştırmalar NaiveBayes DecisionTree SklearnClassifier Analiz

K1 = PKK karşıtı  
K2 = PKK karşıtı ancak liberal  
K3 = Liberal, Görüş Bildirmiyor  
K4 = PKK Sempatizanı  
K5 = ...

Kopyala Excel CSV PDF

Ara:

#	User	CleanTweet	Naive Bayes Analiz DO:0.0%					Decision Tree Analiz DO:0.0%					SklearnClassifier Analiz DO:100.0%				
			K1	K2	K3	K4	K5	K1	K2	K3	K4	K5	K1	K2	K3	K4	K5
1	Mr_garko	6 bedroom duplex with pent at apo resettlement N52m	0.13	0	0.394	0.083	0.394	0	0	0	0	0	0	0	0.5	0.0	0.5
2	sawutul_ummah3	Menbiç'de Pkk'nin korvoyuna bombalı saldırın gerçekleştiği rapor ediliyor	0.13	0	0.394	0.083	0.394	0	0	0	0	0	0	0.5	0.0	0.5	
3	hasanyalcin47	1 Cümledeki anlam bozukluğu yılların gazeteci yazanna yakışmıyor 2 Karşı çıkan HDP'ilerin hiçbirini	0.13	0	0.394	0.083	0.394	0	0	1	0	0	0	0.5	0.0	0.5	

latricti000/canlı-tweet-analizi-tablo/#sklearnclassifier-tablo

Şekil 4.17. Canlı tweet analizi tablo-3 sınıflandırıcı birlikte

Tüm Karşılaştırmalar NaiveBayes DecisionTree SklearnClassifier Analiz

NaiveBayes Doğruluk Oranı:0.0%

K1 = PKK karşıtı  
K2 = PKK karşıtı ancak liberal  
K3 = Liberal, Görüş Bildirmiyor  
K4 = PKK Sempatizanı  
K5 = ...

Kopyala Excel CSV PDF

Ara:

#	User	CleanTweet	RawTweet	Naive Bayes Analiz				
				K1	K2	K3	K4	K5
1	fatma268054	DEMOKRASI PKK'NIN DON LASTİĞİ OLMAMIŞ CHP'YE GÖRE HAKLISINIZ ÇÜNKÜ ARTIK KOL KOLASINIZ ZilletitifikanınKandilAşk	DEMOKRASI PKK'NIN DON LASTİĞİ OLMAMIŞ CHP'YE GÖRE HAKLISINIZ ÇÜNKÜ ARTIK KOL KOLASINIZ #ZilletitifikanınKandilAşk https://t.co/mRzeDnJGB	0.13	0	0.394	0.083	0.394
2	melihimc	Apo'nun İti Olmuş Bir CHP İle Seçimlere Gitmek Atatürk'ün Çocuklarının Zoruna Gidiyor mu Acaba ZilletitifikanınKandilAşk	Apo'nun İti Olmuş Bir CHP İle Seçimlere Gitmek Atatürk'ün Çocuklarının Zoruna Gidiyor mu Acaba #ZilletitifikanınKandilAşk	0.13	0	0.394	0.083	0.394
3	ak_cebe	Chp li Engin Altay' in bu iş 31 Mart'a kalmayacak dediği iş bu o zaman Ancak devlet gerekli tedbirleri ve önlemleri alacaktır Uyanık olmak zorundayız ve unutmayalım ki KORKMA diye başlayan bir İstiklal Margımız var bizim	Chp li Engin Altay' in bu iş 31 Mart'a kalmayacak dediği iş bu o zaman... Ancak devlet gerekli tedbirleri ve önlemleri alacaktır. Uyanık olmak zorundayız ve unutmayalım ki KORKMA diye başlayan bir İstiklal Margımız var bizim.	0.13	0	0.394	0.083	0.394

Şekil 4.18. Canlı tweet analizi tablo-naive bayes sınıflandırıcı

Tüm Karşılaştırmalar NaiveBayes DecisionTree SklearnClassifier Analiz

NLTK-Decision Tree(Karar Ağacı) Analizi Doğruluk Oranı:0.0 %

K1 =PKK karşıtı  
K2 =PKK karşıtı ancak liberal  
K3 =liberal , Görüş Bildirmiyor  
K4 =PKK Sempatizani  
K5 =

Kopyala Excel CSV PDF

Ara:

#	User	CleanTweet	RawTweet	Decision Tree(Karar Ağacı) Analizi				
				K1	K2	K3	K4	K5
1	fatma268054	DEMOKRASİ PKK'NIN DON LASTİĞİ OLMAMKIŞ CHP'YE GÖRE HAKLISINIZ ÇÜNKÜ ARTIK KOL KOLASINIZ #ZilletitfakınınKandiAşkı	DEMOKRASİ PKK'NIN DON LASTİĞİ OLMAMKIŞ CHP'YE GÖRE HAKLISINIZ ÇÜNKÜ ARTIK KOL KOLASINIZ #ZilletitfakınınKandiAşkı https://t.co/mRzeDnJDGB	0	0	0	0	0
2	melihimc	Apo'nun İti Olmuş Bir CHP İle Seçimlere Gitmek Atatürk'ün Çocuklarının Zoruna Gidiyor mu Acaba #ZilletitfakınınKandiAşkı	Apo'nun İti Olmuş Bir CHP İle Seçimlere Gitmek Atatürk'ün Çocuklarının Zoruna Gidiyor mu Acaba ?#ZilletitfakınınKandiAşkı	0	0	0	0	0
3	ak_cebe	Chp İi Engin Altay'ın bu iş 31 Mart'a kalmayacak dediği iş bu o zaman Ancak devlet gerekli tedbirleri ve önlemleri alacaktır Uyanık olmak zorundayız ve unutmayalım ki KORKMA diye başlayan bir İstiklal Marşımız var bizim	Chp İi Engin Altay'ın bu iş 31 Mart'a kalmayacak dediği iş bu o zaman. Ancak devlet gerekli tedbirleri ve önlemleri alacaktır. Uyanık olmak zorundayız ve unutmayalım ki KORKMA diye başlayan bir İstiklal Marşımız var bizim.	0	0	0	0	0
4	MilletinVekili	Ey Muhafazakâr Ülkücü Kardeşim Halen Bu Kararsızlığın Neden	Ey! Muhafazakâr Ülkücü Kardeşim; Halen Bu Kararsızlığın Neden? Kurulan	0	0	0	0	0

Şekil 4.19. Canlı tweet analizi tablo-karar ağacı sınıflandırıcı

Tüm Karşılaştırmalar NaiveBayes DecisionTree SklearnClassifier Analiz

SklearnClassifier Analiz Doğruluk Oranı:100.0 %

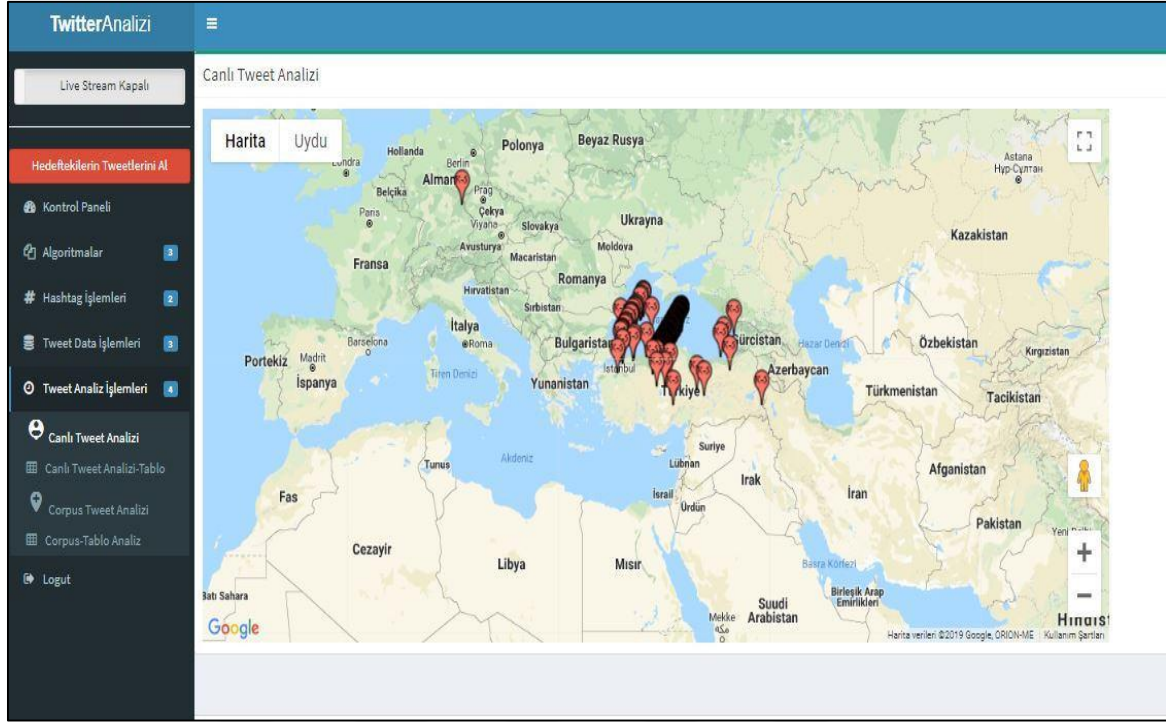
K1 =PKK karşıtı  
K2 =PKK karşıtı ancak liberal  
K3 =liberal , Görüş Bildirmiyor  
K4 =PKK Sempatizani  
K5 =

Kopyala Excel CSV PDF

Ara:

#	User	CleanTweet	RawTweet	NLTK-SklearnClassifier Analizi				
				K1	K2	K3	K4	K5
1	fatma268054	DEMOKRASİ PKK'NIN DON LASTİĞİ OLMAMKIŞ CHP'YE GÖRE HAKLISINIZ ÇÜNKÜ ARTIK KOL KOLASINIZ #ZilletitfakınınKandiAşkı	DEMOKRASİ PKK'NIN DON LASTİĞİ OLMAMKIŞ CHP'YE GÖRE HAKLISINIZ ÇÜNKÜ ARTIK KOL KOLASINIZ #ZilletitfakınınKandiAşkı https://t.co/mRzeDnJDGB	0.0	0	0.5	0.0	0.5
2	melihimc	Apo'nun İti Olmuş Bir CHP İle Seçimlere Gitmek Atatürk'ün Çocuklarının Zoruna Gidiyor mu Acaba #ZilletitfakınınKandiAşkı	Apo'nun İti Olmuş Bir CHP İle Seçimlere Gitmek Atatürk'ün Çocuklarının Zoruna Gidiyor mu Acaba ?#ZilletitfakınınKandiAşkı	0.0	0	0.5	0.0	0.5
3	ak_cebe	Chp İi Engin Altay'ın bu iş 31 Mart'a kalmayacak dediği iş bu o zaman Ancak devlet gerekli tedbirleri ve önlemleri alacaktır Uyanık olmak zorundayız ve unutmayalım ki KORKMA diye başlayan bir İstiklal Marşımız var bizim	Chp İi Engin Altay'ın bu iş 31 Mart'a kalmayacak dediği iş bu o zaman. Ancak devlet gerekli tedbirleri ve önlemleri alacaktır. Uyanık olmak zorundayız ve unutmayalım ki KORKMA diye başlayan bir İstiklal Marşımız var bizim.	0.0	0	0.5	0.0	0.5

Şekil 4.20. Canlı tweet analizi tablo-sklearn bernoulli (N-gram) classifier



Şekil 4.21. Canlı tweet analizi harita üzerinde gösterim

Canlı tweet analizinde harita üzerinde pin adı verilen ve her etiket için ayrı ayrı adlandırılmış olan K1-...K5 arası etiketlerin üzerine gelindiğine o kişinin kim olduğu bilgisi ortaya çıkmaktadır. Son 100 tweet kişi tabanlı ön bir sınıflandırmaya tabi tutulduktan sonra harita üzerinde gösterim Naive Bayes algoritmasına göre yapılmaktadır.

#### 4.6. Corpus Tweet Analizi

Corpus tweet analizi Şekil 4.22, Şekil 4.23, Şekil 4.24 ve Şekil 4.25'deki gibi sistemde kayıtlı tweetlerin %70'i eğitime %30'u test Dataset olacak şekilde ayrıldıktan sonra mümkün olmaktadır. Kullanılan sınıflandırıcılar canlı tweet analizi ile aynıdır. Görünüş olarak aynı ara yüze ve sisteme sahip olmakla birlikte canlı tweet analizinden farkı Corpusu'un rastgele eğitime ve test Dataset'lerine ayrılmasından sonra gerekli analizlerin yapılmasıdır. Ekran görüntüleri aşağıda verilmiştir.

Tüm Karşılaştırmalar NaiveBayes DecisionTree SklearnClassifier Analiz

K1 =PKK karşıtı  
K2 =PKK karşıtı ancak liberal  
K3 =Liberal , Görüş Bildirmiyor  
K4 =PKK Sempatizani  
K5 =

Kopyala Excel CSV PDF Ara:

#	User	CleanTweet	Naive Bayes Analiz DO:75.0 %					Decision Tree Analiz DO:75.0 %					SklearnClassifier Analiz DO:75.0 %					
			K1	K2	K3	K4	K5	K1	K2	K3	K4	K5	K1	K2	K3	K4	K5	
1	75_BOSE	Bu bir 2 One Minute vakasıdır Avrupanın göbeğinde Avrupalı bakanın yüzüne söylendi bu söz "PKK'lıları bu kadar çok seviyorsanız alır götürürsünüz"	0	0	0.439	0.122	0.439	1	0	0	0	0	0	0	0	0.5	0.0	0.5
2	golgenin_Gucu	foxtv kadrosu PKK'ya teşekkür edip rahmet dileyen Karısını öldürseye döven Bağörtülüler Arabistan a diyen foxtv'yi savunanlar Sözde ulusalcı Sözde kadın hakları savunucusu Sözde özgürlükçü Bizde inandık he	0	0	0.701	0.065	0.234	0	0	1	0	0	0	0	0.689	0.001	0.31	
3	akmedya06	Bakin ABD kimleri besliyor da üzerimize salıyor Komünist çeteler PKK YPG ile işbirliği içindeler Suriye gibi olan bir Türkiye'den hortumlanacak olan zengin kaynaklar	0	0	0.701	0.065	0.234	0	0	1	0	0	0	0	0.689	0.001	0.31	

Şekil 4.22. Corpus tweet analizi- 3 sınıflandırıcı tek tablo

Tüm Karşılaştırmalar NaiveBayes DecisionTree SklearnClassifier Analiz

NaiveBayes Doğruluk Oranı:75.0 %

K1 =PKK karşıtı  
K2 =PKK karşıtı ancak liberal  
K3 =Liberal , Görüş Bildirmiyor  
K4 =PKK Sempatizani  
K5 =

Kopyala Excel CSV PDF Ara:

#	User	CleanTweet	RawTweet	Naive Bayes Analiz				
				K1	K2	K3	K4	K5
1	75_BOSE	Bu bir 2 One Minute vakasıdır Avrupanın göbeğinde Avrupalı bakanın yüzüne söylendi bu söz "PKK'lıları bu kadar çok seviyorsanız alır götürürsünüz"	Bu bir 2. One Minute vakasıdır Avrupanın göbeğinde , Avrupalı bakanın yüzüne söylendi bu söz! "PKK'lıları bu kadar çok seviyorsanız alır götürürsünüz" https://t.co/AhceQLe4Pt	0	0	0.439	0.122	0.439
2	golgenin_Gucu	foxtv kadrosu PKK'ya teşekkür edip rahmet dileyen Karısını öldürseye döven Bağörtülüler Arabistan a diyen foxtv'yi savunanlar Sözde ulusalcı Sözde kadın hakları savunucusu Sözde özgürlükçü Bizde inandık he	#foxtv kadrosu: • PKK'ya teşekkür edip, rahmet dileyen @fatihportakal. • Karısını öldürseye döven @Kucukkayalsmail • Bağörtülüler Arabistan'a diyen @DenizCkr #foxtv'yi savunanlar: ✓ Sözde ulusalcı ✓ Sözde kadın hakları savunucusu ✓ Sözde özgürlükçü Bizde inandık he! 🙄🙄 https://t.co/u0pqZFhr1	0	0	0.701	0.065	0.234
3	akmedya06	Bakin ABD kimleri besliyor da üzerimize salıyor Komünist çeteler PKK YPG ile işbirliği içindeler Suriye gibi olan bir Türkiye'den hortumlanacak olan zengin kaynaklar ABD'yi 500 sene daha süper güç yapar	Bakin ABD kimleri besliyor da üzerimize salıyor.. Komünist çeteler PKKYPG ile işbirliği içindeler.. Suriye gibi olan bir Türkiye'den hortumlanacak olan zengin kaynaklar ABD'yi 500 sene daha süper güç yapar. 🙄 https://t.co/qsZFh3hcsh	0	0	0.701	0.065	0.234

Şekil 4.23. Corpus tweet analizi-naive bayes sınıflandırıcı



Tüm Karşılaştırmalar NaiveBayes DecisionTree SklearnClassifier Analiz

NLTK-Decision Tree(Karar Ağacı) Analizi Doğruluk Oranı:0.0 %

K1 =PKK karşıtı  
K2 =PKK karşıtı ancak liberal  
K3 =Liberal , Görüş Bildirmiyor  
K4 =PKK Sempatizani  
K5 =

Kopyala Excel CSV PDF Ara:

#	User	CleanTweet	RawTweet	Decision Tree(Karar Ağacı) Analizi				
				K1	K2	K3	K4	K5
1	fatma268054	DEMOKRASİ PKK'NIN DON LASTİĞİ OLMAMAK İŞİ CHP'YE GÖRE HAKLISINIZ ÇÜNKÜ ARTIK KOL KOLASINIZ ZilletitifikanınKandilAşkı	DEMOKRASİ PKK'NIN DON LASTİĞİ OLMAMAK İŞİ CHP'YE GÖRE HAKLISINIZ ÇÜNKÜ ARTIK KOL KOLASINIZ #ZilletitifikanınKandilAşkı https://t.co/mRzeDnJDGB	0	0	0	0	0
2	melihimc	Apo nun İti Olmuş Bir CHP İle Seçimlere Gitmek Atatürk ün Çocuklarının Zoruna Gidiyor mu Acaba ZilletitifikanınKandilAşkı	Apo'nun İti Olmuş Bir CHP İle Seçimlere Gitmek Atatürk'ün Çocuklarının Zoruna Gidiyor mu Acaba ?#ZilletitifikanınKandilAşkı	0	0	0	0	0
3	ak_cebe	Chp li Engin Altay' in bu iş 31 Mart'a kalmayacak dediği iş bu o zaman Ancak devlet gerekli tedbirleri ve önlemleri alacaktır Uyanık olmak zorundayız ve unutmayalım ki KORKMA diye başlayan bir İstiklal Marşımız var bizim	Chp li Engin Altay' in bu iş 31 Mart'a kalmayacak dediği iş bu o zaman. Ancak devlet gerekli tedbirleri ve önlemleri alacaktır. Uyanık olmak zorundayız ve unutmayalım ki KORKMA diye başlayan bir İstiklal Marşımız var bizim.	0	0	0	0	0
4	MilletinVekili	Ey Muhafazak r Ülkücü Kardeşim Halen Bu Kararsızlığın Neden	Ey! Muhafazakâr Ülkücü Kardeşim; Halen Bu Kararsızlığın Neden? Kurulan	0	0	0	0	0

Şekil 4.24. Corpus tweet analizi- karar ağacı sınıflandırıcı

Tüm Karşılaştırmalar NaiveBayes DecisionTree SklearnClassifier Analiz

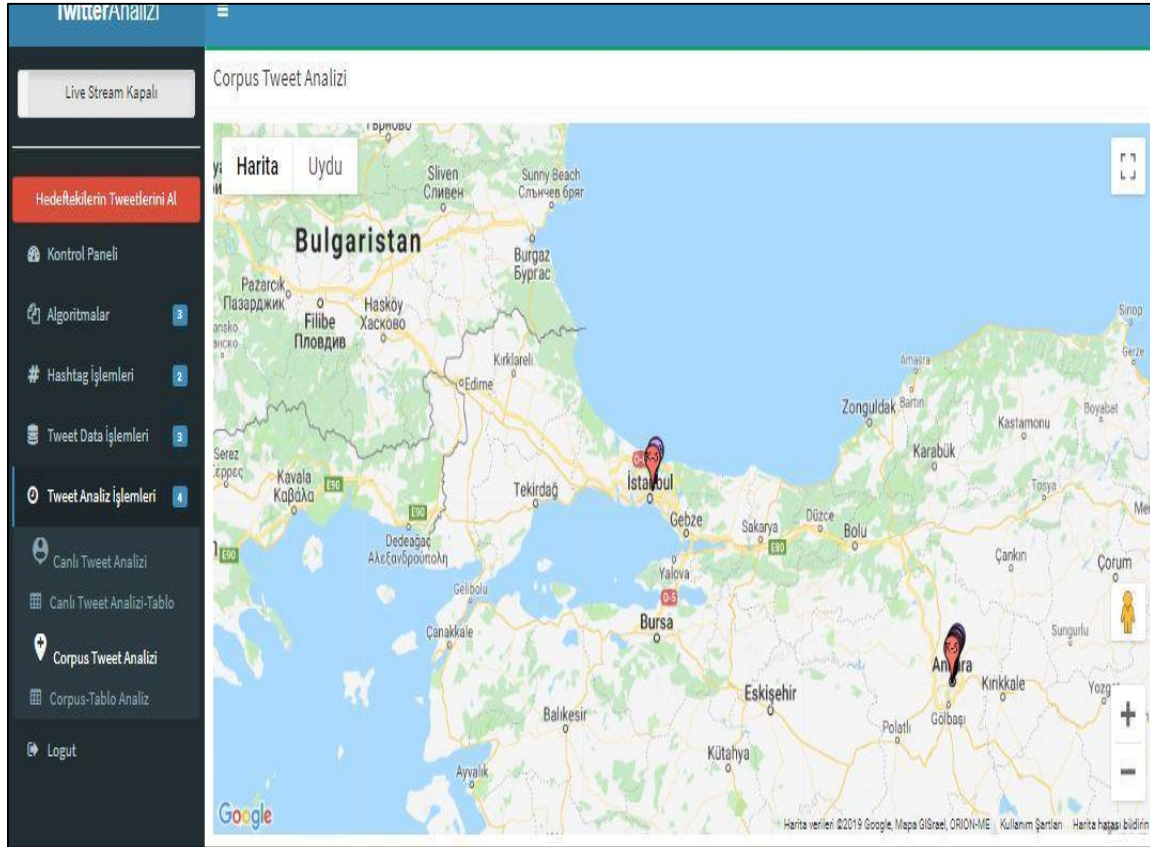
SklearnClassifier Analiz Doğruluk Oranı:75.0 %

K1 =PKK karşıtı  
K2 =PKK karşıtı ancak liberal  
K3 =Liberal , Görüş Bildirmiyor  
K4 =PKK Sempatizani  
K5 =

Kopyala Excel CSV PDF Ara:

#	User	CleanTweet	RawTweet	NLTK-SklearnClassifier Analizi				
				K1	K2	K3	K4	K5
1	75_BOSE	Bu bir 2 One Minute vakasıdır Avrupanın göbeğinde Avrupalı bakanın yüzüne söylendi bu söz "PKK"lıları bu kadar çok seviyorsanız alır götürürsünüz"	Bu bir 2. One Minute vakasıdır! Avrupanın göbeğinde , Avrupalı bakanın yüzüne söylendi bu söz! "PKK"lıları bu kadar çok seviyorsanız alır götürürsünüz" https://t.co/AhceQL4Pt	0	0	0.5	0.0	0.5
2	golgenin_Gucu	foxtv kadrosu PKK'ya teşekkür edip rahmet dileyen Karsını öldüresiye döven Başörtülüler Arabistan a diyen foxtv'yi savunular Sözde ulusalcı Sözde kadın hakları savunucusu Sözde özgürlükçü Bizde inandık he	#foxtv kadrosu: ♦ PKK'ya teşekkür edip, rahmet dileyen @fatihportakal ♦ Karsını öldüresiye döven @Kucukkayalsmail ♦ Başörtülüler Arabistan'a diyen @DenizCkr #foxtv'yi savunular: ✓ Sözde ulusalcı ✓ Sözde kadın hakları savunucusu ✓ Sözde özgürlükçü Bizde inandık he! 🤔🤔 https://t.co/u0pqzFhrs1	0	0	0.689	0.001	0.31
3	akmedya06	Bakin ABD kimleri besliyor da üzerimize salıyor Komünist çeteler PKK YPG ile işbirliği içindeler Suriye gibi olan bir Türkiye'den hortumlanacak olan zengin kaynaklar ABD'yi 500 sene daha süper güç yapar	Bakin ABD kimleri besliyor da üzerimize salıyor.. Komünist çeteler PKK/YPG ile işbirliği içindeler.. Suriye gibi olan bir Türkiye'den hortumlanacak olan zengin kaynaklar ABD'yi 500 sene daha süper güç yapar.. 🤔 https://t.co/qsZfH3hcsh	0	0	0.689	0.001	0.31

Şekil 4.25. Corpus tweet analizi- sklearn bernoulli (n-gram) classifier



Şekil 4.26. Corpus tweet analizi-harita görüntüsü

Bu çalışmada elde edilen sonuçlar göstermiştir ki kullanılan 3 algorithmadan sadece Karar Ağacı algoritması Çizelge 4.1’de görüldüğü gibi en yüksek oranda başarı sağlamıştır. Yalnız tek başına bu algoritmayı kullanarak çıkarım elde etmek fayda sağlamamaktadır. Bu yüzden sonuç olarak bu üç algoritmanın hepsini aynı anda çalıştırıp ve çıkan sonuçları harmanlayıp nihai bir sonuç elde etmek gerekmektedir. Çalışma %99.24 oranında kişilerin geriye dönük sosyal medya paylaşımlarını izleyerek doğal dil işleme tekniğini kullanarak ileride hangi örgütlere sempati duyup duymayacakları veya hangi örgüte mensup olup olmadıkları hakkında bize büyük oranda katkı sağlamaktadır.

Çizelge 4.1. Algoritmaların başarı oranı

Algoritma Adı	Başarı oranı
Navi Bayes	%96.21
Karar Ağacı	%99.24
N-Gram	%88.7

Bu tespitler %100 doğru olmayabilir, Bu da doğaldır, çünkü bilgisayarlar insan beyni gibi çalışmadığı için ironi, şaka ve gönderme gibi kavramları anlamakta zorluk çekmektedirler. Bir kişi sırf bir paylaşımda ironi yapmış olabilir ama makine onun ironi yapmış olduğunu anlayamaz ve onu gerçek bir metin sanarak terör örgütü listesine sokabilir. Eğer hedefteki kişi ironi yapmaya veya dolaylı konuşmaya devam ederse sistem onu yanlışlıkla her hangi bir terör örgütü mensubu veya terör örgütü sempatzanı olarak sınıflandırabilir.



## 5. DEĞERLENDİRME VE SONUÇ

Duygusal Twitter Analizi, gelişen teknoloji sayesinde insanların ilişkilerinin siber âleme kayması ve neredeyse tüm duygu ve düşüncelerinin bu platformlarda ifade etmelerinin bir sonucu olarak bu ortamda paylaşılan içeriklerden anlamlı sonuçlar çıkarma amacını taşımaktadır. Sosyal Medya araçlarından neredeyse en hızlı ve etkileme açısından en etkili olan Twitter model alınmıştır. Twitter içeriği olan tweetlerden Makine Öğrenimi ve Yapay Zekâ Algoritmaları kullanılarak bir Corpus oluşturulmuştur. Burada yapılması gereken algoritmalar ve bilgisayarların işlem gücü kullanılarak anlamlı sonuçlar çıkarmadır. Önceden oluşturulan Corpus, makinenin etiketlerden yola çıkarak kendisine sorulan içeriklerin ne olduğunu sınıflandırmaya yaramaktadır. Elde edilen sonuçlar bu şekilde bir analizin mümkün olacağını ortaya koymaktadır. Duygusal Twitter Analizinin amacı her ne kadar kullanıcıların paylaşımlarındaki duygu tiplerinin ortaya çıkarılması olsa dahi, ilerleyen teknik ve alt yapıyla, sosyal medya ortamında işlenen suçlar, çocuk pornografisi, teröre destek ve eleman kazanma gibi birçok Kriminal vakaların çözüme kavuşturulması için hem politikacılara hem de bu işin uzmanlarına gerekli desteği verecektir.

Bu çalışmada ilk başta .NET teknolojisi kullanılmak istenmiştir. Yalnız Microsoft firması makine öğrenme alanında çalışmanın yapıldığı esnada tam verimli çalışan kütüphanelerin olmaması sebebiyle 2. sıraya yerleşmiştir. Ayrıca Java diline de bir bakış yapılmıştır. Ve aynı sorun Java dilinde de görülmüştür. En son, geniş kütüphane ve geliştirici camiasına sahip olan Python dili seçilmiştir. Yalnız Python dili başlı başına Web tabanlı uygulamaları desteklememektedir. Bu yüzden Django Framework web tabanlı yazılım geliştirmek için uygun görülmüştür. Bu Framework ne kadar Python dilinde web uygulama geliştirmek için yapılmış olsa da büyük zorluklar çıkarmıştır. Zorlukların başında güncelleme sorunu gelmektedir. Bu Framework her güncellendiğinde o güncellemeyi alıp bu projeye uygulamak çok uzun zamanlar alabilmektedir.

Diğer taraftan Twitter firması son zamanlarda bu tarz çalışmaların sayısının artmasından kaynaklı Twitter sorgulama web Api'lerinde kısıtlamaya gitmiştir. İlk başlarda sınırsız olarak sorgu yapma hakkı tanıyan Twitter, daha sonra sadece 200 tweeti ücretsiz olarak sorgulama hakkı tanımıştır. Bu da bu çalışmanın asıl amacı olan bir kişinin geçmişe dönük duyguları ve terör örgütüne yakınlığının anlaşılması için büyük bir sıkıntı teşkil etmiştir. Tabii ki Twitter bu soruna bir çözüm getirmektedir. Twitter ücret karşılığı sınırsız sorgulama



hizmeti sunmakta, lakin bu çalışma kapsamında maddi kaynakların kısıtlı olmasından dolayı bu ücretli web Api kullanılmamıştır.

Genel olarak bu çalışma başta ifade edildiği gibi yapay zekâ algoritmalarıyla sosyal medyada ikilerin siyasi eğilimlerini tespit edebilme tezi ispatlamıştır. Yapay Zekâ teknikleri kullanılarak insanların davranış biçimleri sınıflandırılabilir ve düşünce tarzları anlaşılabilir. Hatta bir adım ileri atarak gelecekteki davranışları bile tahmin edilebilir. Ayrıca Akıllı telefonlarda kullanılan Twitter tarzı sosyal medya bu kişilerin konumsal verilerini de sunmaktadır. Kişilerin paylaşımları konumsal verilerini büyük veri kapsamında işledikten sonra her bir ferдин bir toplumda hangi düşünceye yakın olduğunu öngörebilir ve hatta kişinin nerede yaşadığı, hangi konumda yaşadığı bile görülebilir. Peki, bu nasıl bir yarar sağlar? Örneğin bir ülkenin aşırı ve radikal bir terör örgütünden mustarip olduğunu varsayalım. Bu ülkenin emniyet birimleri bu köktenci grup üyelerinin nerede yaşadığı ve daha çok ülkenin hangi kısımlarında dağılımda olduklarını çıkarmak için belki yıllarını harcayacaktır. Ama bu çalışma vasıtasıyla sadece sosyal medyadan yola çıkarak en azından bu emniyet birimlerine bir genel fikir sunabilir. Belki bu çalışmada sadece Ankara şehri temel alınmıştır ama bir de Türkiye'yi veya herhangi bir ülke düşünüldüğünde, bir ülkede köktenci grupların hangi bölgelerde, örneğin kuzeybatıda mı sahil şeridinde mi yoksa ülkenin ortasında mı olduğuna dair kanaat getirilebilir. Ayrıca daha Eylem yapmamış ve herhangi bir sosyal veya toplumsal olaya karışmamış kişiler bile daha eyleme geçmeden önceden tespit edilebilir ve önlenir.

Yapılan bu çalışmada 14048 Tweet girdisi kullanılmıştır. Bu 14048 tweetden %70'i sistemin eğitimi için ve geri kalan %30'u ise sistemin testi için kullanılmıştır. Bu verileri sayesinde Karar Ağacı algoritması en çok verimi sağlamıştır. Diğer algoritmalar arasından Navie Bayes %96.21'lik bir başarı oranıyla ikinci sırada durmaktadır. En az verim alınan algoritma %88.7'lik oranla N-Gram algoritması olmuştur. Bu çalışmada en yüksek başarıyı Navie Bayes algoritması elde etmiş olsa da tek başına amaçlanan hedefe varmak için yeterli değildir. Bir kişinin bir terör örgütüne yakınlık duyup duymadığı sadece bir algoritmanın çıkarımına bırakılamaz. Bu yüzden kullanılan her üç algoritmanın sonuçlarını karma bir şekilde kullanılması lazım. Görüldüğü üzere kullanılan üç algoritmanın nihayi sonuçlarını birleştirerek haritada yüksek doğruluk oranında sonuç elde edilmiştir. Son olarak bu çalışmada geliştirilen uygulama sadece bir karar destek mekanizması niteliği

tařımaktadır. Bu uygulama tek başına kişileri tespit etse de tek başına karar verme yeteneğine sahip değildir. Nihai karar bir insan tarafından verilmelidir.





## KAYNAKLAR

1. İnternet: Global Digital 2018 Raporu. (2018). URL: <https://wearesocial.com/blog/2018/01/global-digital-report-2018>, Son Erişim Tarihi: 22.02.2019.
2. Pak, A. and Paroubek, P. (2010). *Twitter as a Corpus for Sentiment Analysis and Opinion Mining*, Universit'e de Paris-Sud, Laboratoire LIMSI-CNRS, Batiment, 10, 1320-1326.
3. Nagar, R., Yuan, Q., Freifeld, C. C., Santillana, M., Nojima, A., Chunara, R. and Brownstein, J. S. (2014). A case study of the New York City 2012-2013 influenza season with daily geocoded Twitter data from temporal and spatiotemporal perspectives. *Journal of Medical Internet Research*, 16(10), e236.
4. Wang, H., Can, D., Kazemzadeh, A., Bar, F. and Narayanan, S. (2012, 8-14 July). *A System for Real-time Twitter sentiment analysis of 2012 U.S. Presidential Election Cycle*, Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, Jeju, Republic of Korea 115-120,
5. Yıldırım, A. (2014). Bir halkla ilişkiler aracı olarak Twitter: T.C. Sağlık Bakanlığı örnek incelemesi, *Gümüşhane Üniversitesi İletişim Fakültesi Elektronik Dergisi*, 2(4), 234-253.
6. Kwak, H., Lee, C., Hosung P., Moon, S. (2010). *What is twitter, a social network or a news media?*, International World Wide Web Conference Committee (IW3C2), North Carolina, USA.
7. Kuyucu, M. ve Karahisar, T. (2013). *Yeni iletişim teknolojileri ve yeni medya*, İstanbul: Zinde Reklam Yayıncılık.
8. İnternet: Bakshy, E., Hofman, J. M., A. Mason, W. and J. Watts, D. (2011). Everyone's an Influencer: Quantifying Influence on Twitter, WSDM'11, China, URL: <http://misc.si.umich.edu/media/papers/wsdm333wbakshy.pdf>, Son Erişim Tarihi: 13.04.2019.
9. Bayraktutan, G., Binark, M., Çomu, T., Burak, D., İslamoğlu, G. ve Telli A. (2014). Siyasal iletişim sürecinde sosyal medya ve Türkiye'de 2011 seçimlerinde Twitter kullanımı, *Türk Dünyası Sosyal Bilimler Dergisi*, 68, 59-96.
10. Karaçor, S. (2009). Yeni iletişim teknolojileri, siyasal katılım, demokrasi, *Yönetim ve Ekonomi Dergisi*, 16(2), 121-131.
11. Şeker, Ş. E. (2013). *İş zekası ve veri madenciliği*. İstanbul: Cinius Yayınları, 12-16.
12. Prytherch, R. (2005). *Harrod's librarians' glossary and reference book : A dictionary of over 10, 200 Terms*. 10. Hampshire: Ashgate Publishing Limited, 31-32.
13. Yılmaz, M. (2009). Enformasyon ve Bilgi Kavramları Bağlamında Enformasyon Yönetimi ve Bilgi Yönetimi. *Ankara Üniversitesi Dil ve Tarih-Coğrafya Fakültesi Dergisi* 49(1), 95-118.

14. Jeffery, K. (2016). Data is the New Oil. *Best Practices for Data Management & Sharing. Dü. The Joint Research Centre (JRC)*. Ispra, Italy.
15. İnternet: Büyük Veri Tanımı, Gartner. URL: <https://www.gartner.com/it-glossary/big-data/>, Son Erişim Tarihi: 22.02.2019.
16. İnternet: Gartner's Big Data Definition Consists of Three Parts, Not to Be Confused with Three "V"s, URL: <https://www.forbes.com/sites/gartnergroup/2013/03/27/gartners-big-data-definition-consists-of-three-parts-not-to-be-confused-with-three-vs/#7ec4647a42f6>, Son Erişim Tarihi: 22.02.2019.
17. İnternet: Laney D., Application Delivery Strategies. (2012). <https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>, Son Erişim Tarihi: 22.02.2019.
18. Demchenko, Y., Zhao, Z., Grosso, P., Wibisono, A. (2012). *de Laat C: Addressing big data challenges for scientific data infrastructure*. In IEEE 4th International Conference on Cloud Computing Technology and Science (CloudCom 2012). Taipei, Taiwan: IEEE Computing Society, USA, 614–617.
19. İnternet: URL: Lesk Micheal, How Much Information Is There I he World?, <https://courses.cs.washington.edu/courses/cse590s/03au/lesk.pdf>, Son Erişim Tarihi: 15.03.2019.
20. Jun Lee, S. and Siau, K. (2001). A review of data mining techniques. *Industrial Management & Data Systems*, 101(1), 41-46.
21. Can, Ü. and Alataş, B. (2017). Review of sentiment analysis and opinion mining algorithms. *International Journal of Pure and Applied Sciences*, 3(1), 75-111.
22. Orbanz, P., and Teh, Y. W. (2010). Bayesian nonparametric models. *Encyclopedia of Machine Learning*, 81-89.
23. Nicolosi, N. (2008). *Feature selection methods for text classification*. Department of Computer Science, Rochester Institute of Technology, Technical Report.
24. Gastaldo, P., Bisio, F., Gianoglio, C., Ragusa, E. and Zunino, R. (2017). Learning with similarity functions: a novel design for the extreme learning machine. *Neurocomputing*, 261, 37-49.
25. Silva, A. P. D. (2017). Optimization approaches to supervised classification. *European Journal of Operational Research*, 261(2), 772-788.
26. Mazzeo, G. M., Masciari, E. and Zaniolo, C. (2017). A fast and accurate algorithm for unsupervised clustering around centroids. *Information Sciences*, 400, 63-90.
27. Drugan, M. M. (2019). Reinforcement learning versus evolutionary computation: A survey on hybrid algorithms. *Swarm and Evolutionary Computation*, 44, 228-246.
28. Sengupta, N., and Sil, J. (2012, August). *Comparison of supervised learning and reinforcement learning in intrusion domain*. In International Conference on Information Processing. Springer, Berlin, Heidelberg, 546-551.

29. Ding, S., Jia, H., Du, M. and Xue, Y. (2018). A semi-supervised approximate spectral clustering algorithm based on HMRF model. *Information Sciences*, 429, 215-228.
30. Copeland B. J. (2004). *The essential turing: Seminal writings in computing, logic, philosophy, artificial intelligence, and artificial life: Plus the secret of enigma*, Oxford University Press, 353-361.
31. Pirim, H. (2006). Yapay zeka, *Journal of Yasar University*, 1(1), 81-93.
32. Adalı, E. (2017). Yapay zeka, *İstanbul Teknik Üniversitesi Vakfı Dergisi*, 75, 8-14.
33. Patan, İ. (2017). *Siyasal halkla ilişkilerde yeni iletişim teknolojilerinin kullanımı, TBMM'de temsil edilen siyasi partilerin 2015 genel seçimlerinde Facebook ve Twitter kullanımları üzerine bir inceleme*, Yayınlanmamış Yüksek Lisans Tezi, Giresun Üniversitesi, Sosyal Bilimler Enstitüsü, Giresun.
34. İnternet: URL: Twitter For Developers. <https://developer.Twitter.com/>, Son Erişim Tarihi: 12.03.2019.
35. İnternet: Yüksektepe, M. (2016). Django için Türkçe Kılavuz, URL: [www.muslu.org](http://www.muslu.org), Son Erişim Tarihi: 21.07.2018.
36. İnternet: URL: <https://docs.djangoproject.com/en/dev/faq/general/#django-appears-to-be-a-mvc-framework-but-you-call-the-controller-the-view-and-the-view-the-template-how-come-you-don-t-use-the-standard-names>, Son Erişim Tarihi: 13.03.2018.
37. İnternet: Tweepy. URL: <https://Tweepy.readthedocs.io/en/v3.5.0/>, Son Erişim Tarihi: 13.03.2018.
38. Karabulut, Y. E. ve Küçüksille, E. U. (2018). Twitter profesyonel izleme ve analiz aracı, *Teknik Bilimleri Dergisi*, 8(2), 17-24.
39. İnternet: Twitter Documentation, URL: <https://developer.twitter.com/en/docs.html>, Son Erişim Tarihi: 2014.
40. İnternet: PotgreSql. URL: <https://www.postgresql.org/>, Son Erişim Tarihi: 13.03.2018.
41. Tubitak, (2018). *PostgreSQL dünyanın en gelişmiş açık kaynaklı veritabanı*, Yazılım Teknolojileri Araştırma Enstitüsü, Ankara.
42. İnternet: Yazıcı, V. (2006). *PostgreSQL ile Programlama*, URL: <http://www.students.itu.edu.tr/~yazicivo/>, Son Erişim Tarihi: 19.01.2019.
43. İnternet: GoogleMapApi <https://cloud.google.com/maps-platform/>, Son Erişim Tarihi: 13.03.2018.
44. İneç, Z. F. ve Akpınar, E. (2011). Web haritalama hizmetleri (WMS) uygulamalarının teknik ve performans bakımından incelenmesi, *Marmara Coğrafya Dergisi*, 24, 403-4332.
45. İnternet: URL: <https://www.wowebook.info/book/jquery-novice-to-ninja/>, Son Erişim Tarihi: 15.03.2018.

46. İnternet: URL: <https://mesutd.com/datatables-eklentisi-ile-tablolarda-gelismis-ozelliklerin-kullanimi>, Son Erişim Tarihi: 05.11.2017.
47. Özmutlu, H. C. ve Çağlar, B. (2009). Arama motorlarında yeni konu tanılamada karakter N-Gram ve yapay sinir ağları uygulaması, *Uludağ Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi*, 14(2), 15.
48. İnternet: URL: <http://textblob.readthedocs.io/en/dev/index.html>, Son Erişim Tarihi: 13.03.2019.
49. İnternet: URL: <https://www.nltk.org/>, Son Erişim Tarihi: 13.03.2018.
50. İnternet: URL: <https://textblob.readthedocs.io/en/dev/classifiers.html>, Son Erişim Tarihi: 13.03.2019.
51. İnternet: URL: [https://textblob.readthedocs.io/en/dev/api\\_reference.html#api-classifiers](https://textblob.readthedocs.io/en/dev/api_reference.html#api-classifiers), Son Erişim Tarihi: 13.03.2019
52. Javadi, T. ve Barışcı, N. (2018). *Duygusal twitter analizi*, ISAS 2018-Ist International Symposium on Innovative Approaches in Scientific Studies, Kemer-Antalya 2, 380-385.

## ÖZGEÇMİŞ

### Kişisel Bilgiler

Soyadı, adı : JAVADI, Tohid  
 Uyuğu : İran  
 Doğum tarihi ve yeri : 1988, Urmiye  
 Medeni hali : Evli  
 Telefon : 0553-114-8515  
 e-mail : tohid.javadi@gmail.com



### Eğitim

Derece	Eğitim Birimi	Mezuniyet Tarihi
Yüksek Lisans	Gazi Üniversitesi / Bilgisayar Mühendisliği	Devam Ediyor
Lisans	Azad-e Eslami Üniversitesi / Bilgisayar Mühendisliği	2011
Lise	Malik Eshter	2008

### İş Deneyimi

Yıl	Yer	Görev
2014-2016	Desmer Güvenlik	Yazılım Takım Lideri
2016-2014	FastSoft Yazılım AŞ	Yazılım Takım Lideri

### Yabancı Dil

İngilizce, Farsça

### Yayınlar

- Javadi, T. ve Barışcı, N. (2018). *Duygusal twitter analizi*, ISAS 2018-Ist International Symposium on Innovative Approaches in Scientific Studies, Kemer-Antalya 2, 380-385.

### Hobiler

Yüzme, Evcil Hayvan





*GAZİ GELECEKTİR..*